PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-194796
(43)Date of publication of application : 21.07.1999

(51)Int.Cl. G10L 3/02
G10L 7/04
G11B 20/10
H03M 7/30

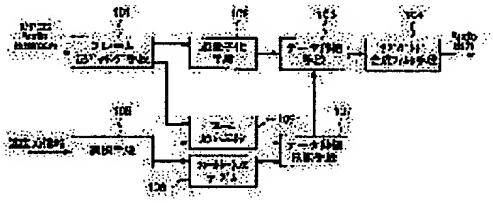
(21)Application number : 10-218925
(22)Date of filing : 03.08.1998

(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD
(72)Inventor : MISAKI MASAYUKI
TANIGUCHI HIROTSUGU
TAGAWA JUNICHI
MATSUMOTO MICHIO

(30)Priority
Priority number : 09300121 Priority date : 31.10.1997 Priority country : JP

(54) SPEECH REPRODUCING DEVICE

(57)Abstract:
PROBLEM TO BE SOLVED: To obtain good quality tone by a simple composition of a speed conversion processing in the case of decoding by the frame.
SOLUTION: This speech reproducing device is provided with a frame inverse packing means 101 for decoding speech signals inputted by the frame, a data expansion and contraction means 103 for having the decoded data in the frame subjected to time base transformation processing, a frame sequence table 108 in which a sequence of an expansion and contraction processing to each frame according to a given speed ratio, a frame count means 106 for counting the number of the frames of the input speed signals, and a data expansion and contraction control means 107 which refers to the frame sequence table 108 based on the count value from the frame count means 106 and indicates the data expansion and contraction means 103 by with method of 'time base compression', 'time base expansion', or 'witout time base transformation' to process the frames, and the data expansion and contraction means 103 processes the speech signal by a time base conversion based on the indication signal from the data expansion and contraction control means.



LEGAL STATUS

[Date of request for examination]	03.08.1998
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3017715
[Date of registration]	24.12.1999

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)特　許　公　報（Ｂ２）　(11)特許番号
特許第3017715号
(P3017715)
(45)発行日　平成12年３月13日(2000.3.13)　(24)登録日　平成11年12月24日(1999.12.24)

(51)Int.Cl.⁷	識別記号	Ｆ Ｉ	
Ｇ１０Ｌ　21/04		Ｇ１０Ｌ　3/02	Ｃ
19/02		Ｇ１１Ｂ　20/02	Ｇ
Ｇ１１Ｂ　20/02		Ｇ１０Ｌ　7/04	Ｇ

請求項の数13(全 30 頁)

(21)出願番号　特願平10－218925
(22)出願日　平成10年８月３日(1998.8.3)
(65)公開番号　特開平11－194796
(43)公開日　平成11年７月21日(1999.7.21)
審査請求日　平成10年８月３日(1998.8.3)
(31)優先権主張番号　特願平9－300121
(32)優先日　平成９年10月31日(1997.10.31)
(33)優先権主張国　日本（ＪＰ）

(73)特許権者　000005821
松下電器産業株式会社
大阪府門真市大字門真1006番地
(72)発明者　三▲さき▼　正之
大阪府門真市大字門真1006番地　松下電器産業株式会社内
(72)発明者　谷口　宏嗣
大阪府門真市大字門真1006番地　松下電器産業株式会社内
(72)発明者　田川　潤一
大阪府門真市大字門真1006番地　松下電器産業株式会社内
(74)代理人　100081813
弁理士　早瀬　憲一

審査官　南　義明

最終頁に続く

(54)【発明の名称】　音声再生装置

(57)【特許請求の範囲】
【請求項１】　音声復号化手段、選択手段、フレームシーケンステーブル、フレームカウント手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、
音声復号化手段は、入力される音声信号をフレーム単位で復号し、
選択手段は、与えられる速度比に対応したフレームシーケンスをフレームシーケンステーブルへ出力すると共に、該フレームシーケンスのフレームサイクル数をフレームカウント手段へ出力し、
フレームシーケンステーブルは、選択手段からのフレームシーケンスを記憶し、
フレームカウント手段は、フレームサイクル数に基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、
データ伸縮制御手段は、フレームカウント手段のカウント値に対応したフレームシーケンステーブルのフレームシーケンスを参照して、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、
データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行う音声再生装置。
【請求項２】　音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された音声信号を復号する請求項１に記載の音声再生装置。
【請求項３】　フレームシーケンスは、連続する時間軸圧縮フレームのフレーム数と、連続する時間軸処理無し

フレームのフレーム数がいずれも最小となるよう配置された請求項１に記載の音声再生装置。

【請求項４】　フレームシーケンスは、連続する時間軸伸長フレームのフレーム数と、連続する時間軸処理無しフレームのフレーム数がいずれも最小となるよう配置された請求項１に記載の音声再生装置。

【請求項５】　音声復号化手段、伸縮頻度制御手段、フレームカウント手段、エネルギー演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、

音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された符号化音声信号を復号し、

伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、

フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、

エネルギー演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギーを推定し、

フレーム選択手段は、フレームサイクル数Ｎｆのフレーム内でエネルギーの小さいフレームからＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、

データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、

データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行う音声再生装置。

【請求項６】　音声復号化手段、伸縮頻度制御手段、フレームカウント手段、定常性演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、

音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された音声信号を復号し、

伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、

フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、

定常性演算手段は、符号化音声信号のスケールファクタ選択情報をもとにフレームサイクル数Ｎｆ分の符号化音声信号の定常性を推定し、

フレーム選択手段は、フレームサイクル数Ｎｆのフレーム内での定常性の高いフレームからＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、

データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、

データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行う音声再生装置。

【請求項７】　音声復号化手段、伸縮頻度制御手段、フレームカウント手段、エネルギー変化度合演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、

音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された音声信号を復号し、

伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、

フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、

エネルギー変化度合演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギー変化度合を推定し、

フレーム選択手段は、フレームサイクル数Ｎｆのフレーム内でエネルギー変化度合に基づき継時マスキング効果による処理劣化が少ないフレームからＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、

データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、

データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行う音声再生装置。

【請求項８】　音声復号化手段、伸縮頻度制御手段、フレームカウント手段、演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、

音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された音声信号を復号し、

伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、

フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレ

ーム数をカウントし、
演算手段は、エネルギー演算手段、定常性演算手段、エネルギー変化度合演算手段のいずれか２つ以上を備え、
エネルギー演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギーを推定し、
定常性演算手段は、符号化音声信号のスケールファクタ選択情報をもとにフレームサイクル数Ｎｆ分の符号化音声信号の定常性を推定し、
エネルギー変化度合演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギー変化度合を推定し、
フレーム選択手段は、演算手段の出力をもとにＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、
データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、
データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行う音声再生装置。
【請求項９】　データ伸縮手段は、クロスフェード手段を備え、
クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを重み付け加算する請求項１～８のいずれかに記載の音声再生装置。
【請求項１０】　データ伸縮手段は、相関演算手段、クロスフェード手段を備え、
相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、セグメント間の相関値を演算し、相関値が高くなる位置で重み付け加算するためのシフト量を決定し、クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算する請求項１～８のいずれかに記載の音声再生装置。
【請求項１１】　音声復号化手段は、符号化音声信号を帯域毎に復号し、
データ伸縮手段は、相関演算手段、帯域毎のクロスフェード手段を備え、
相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、ピッチ周波数を包含する帯域においてセグメント間の相関値を演算し、相関値が高くなる位置で重み付け加算するためのシフト量を決定し、
各クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算する請求項１～８のいずれかに記載の音声再生装置。
【請求項１２】　音声復号化手段は、符号化音声信号を帯域毎に復号し、
データ伸縮手段は、相関演算手段、帯域毎のクロスフェード手段を備え、
相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、平均エネルギーが最大となる帯域においてセグメント間の相関値を演算し、相関値が高くなる位置で重み付け加算するためのシフト量を決定し、
各クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算する請求項１～８のいずれかに記載の音声再生装置。
【請求項１３】　音声復号化手段は、符号化音声信号を帯域毎に復号し、
データ伸縮手段は、相関演算手段、帯域毎のクロスフェード手段を備え、
相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、各帯域においてセグメント間の相関値を演算し、相関値が最大の帯域において相関値が高くなる位置で重み付け加算するためのシフト量を決定し、
各クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算する請求項１～８のいずれかに記載の音声再生装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、音声速度を所望の値に変換して聴取する事が可能な音声再生装置に関する。
【０００２】
【従来の技術】音声を高能率に符号化して、記憶媒体へ蓄積、あるいは通信網を利用して伝送する技術が近年実用化され広く利用されている。
【０００３】このような技術に関し、国際標準規格のMPEG方式を用いて、音声（オーディオ）を再生する装置として、例えば特開平９－７３２９９号公報に開示されているものがある。このMPEGオーディオ再生装置のブロック図を図１９に示す。以下、図１９を参照しながら、従来の音声再生装置について説明する。
【０００４】図１９に示すように、MPEGオーディオ再生

装置１は、再生速度検出回路２、MPEGオーディオデコーダ３、話速変換処理回路４、Ｄ／Ａコンバータ５、オーディオアンプ６から構成されている。さらに話速変換処理回路４は、フレームメモリ３４、話速変換部３５、リングメモリ３２、アップダウンカウンタ３３、読み出しクロック生成回路３６で構成されている。

【０００５】MPEGオーディオ再生装置１には、MPEGオーディオ方式にて符号化されたMPEGオーディオストリームが入力される。MPEGオーディオデコーダ３では、上記MPEGオーディオストリームがディジタル信号のオーディオ出力に復号される。MPEGオーディオの方式およびフォーマットの内容に関しては、現在では種々の文献等に記述されており、例えば「ISO/IEC IS 11172 Part3 : Audio」に記載されている。

【０００６】一方、例えば２倍速、０．５倍速などの速度情報が再生速度検出回路２に入力され、この再生速度検出回路にて速度情報（再生速度）を検出してデコードクロックを生成する。このデコードクロックは話速変換処理回路４およびMPEGオーディオデコーダ３へ供給される。当該MPEGオーディオデコーダ３にてデコードされたオーディオ信号は、話速変換処理回路４に入力され、与えられた上記速度情報に基づき、さらに時間軸圧縮／伸長あるいは無音削除／挿入を施されて、所定の話速変換が行われ、この話速変換された出力がスピーカ２３から再生されることとなる。

【０００７】

【発明が解決しようとする課題】しかしながら、上記のような、所定時間長のフレーム単位でのデコードを行うMPEGオーディオのような符号化方式において、複数フレーム間にまたがるデータ処理を実施する際には、多数のバッファメモリなどが必要かつ処理が複雑となり、ハードウェア構成が大規模となる問題を生じることになる。

【０００８】さらに、同様に、国際標準規格のMPEG方式を用いて、音声（オーディオ）を再生する装置として、特開平９－８１１８９号公報に開示されているものがある。このMPEGオーディオ再生装置のブロック図を、図２０に示す。以下、図２０を参照しながら、従来の音声再生装置について説明する。

【０００９】図２０に示すように、１７０１は、入力される帯域信号１をＴｆサンプル長の１フレーム分、分割し保持する第１のフレーム分割装置、１７０２は、入力される帯域信号２をＴｆサンプル長の１フレーム分、分割し保持する第２のフレーム分割装置、１７０３は、入力される帯域信号３をＴｆサンプル長の１フレーム分、分割し保持する第３のフレーム分割装置、１７０４は、入力される帯域信号４をＴｆサンプル長の１フレーム分、分割し保持する第４のフレーム分割装置である。

【００１０】上記において、入力される帯域信号１～４は、通常の時間軸信号を４帯域に帯域分割するとともに４分の１にダウンサンプリングするようなフィルタバンクによって帯域分割された、それぞれの帯域信号であり、帯域信号１は、最も低域の帯域信号、帯域信号４、は最も高域の帯域信号であるとする。

【００１１】１７１０は、音声のピッチ成分が含まれる帯域の帯域信号の前半の信号と、後半の信号とを、ｎサンプルだけオーバーラップさせた時の該オーバーラップ範囲における両信号間の相関値S(n)を求め、該相関値S(n)が最大値となるn をTcとして検出する相関関数算出装置、１７１１は、聴取者からの再生速度F の指定を検出する再生速度検出装置、１７１２は、相関関数検出範囲に制限を設けるための相関関数検出範囲制御装置、１７０５は、第１のフレーム分割装置１７０１によって分割され保持された帯域信号の前半の信号と、後半の信号とを、Ｔｃサンプル分オーバーラップさせてクロスフェード処理する第１のクロスフェード処理装置、１７０６は、第２のフレーム分割装置１７０２によって分割され保持された帯域信号の前半の信号と、後半の信号とを、Tcサンプル分オーバーラップさせてクロスフェード処理する第２のクロスフェード処理装置、１７０７は、第３のフレーム分割装置１７０３によって分割され保持された帯域信号の前半の信号と、後半の信号とを、Tcサンプル分オーバーラップさせてクロスフェード処理する第３のクロスフェード処理装置、１７０８は、第４のフレーム分割装置１７０８によって分割され保持された帯域信号の前半の信号と、後半の信号とを、Tcサンプル分オーバーラップさせてクロスフェード処理する第４のクロスフェード処理装置、１７０９は、上記クロスフェード処理された４帯域の帯域信号を帯域合成する帯域合成フィルタである。

【００１２】図２１は、音声信号の主要ピッチ成分が含まれる周波数帯域について、その１フレーム分の時間軸波形を表した図である。図２２は、図２１に示した１フレームの信号を、その前半の信号部分と、後半の信号部分との２セグメントに分割して上下に並べた図である。図２３は、図２２における２セグメント間の相関関数を求めた値を示したグラフである。図２４は、相関関数が最大となる時刻に後半の信号成分であるセグメントをずらせた様子を定性的に示した図である。図２５は、２セグメント間をTc時間オーバーラップさせてクロスフェード処理する様子を示した図である。

【００１３】以上のように構成された再生装置について、以下その動作について、図２１から図２５を用いて説明する。まず入力される帯域信号１の１フレーム分（Ｔｆサンプル長）のデータは、図２１に示すように、音声信号の主要ピッチ成分を含んでいるものとする。そして、この１フレーム分のデータは、第１のフレーム分割装置１７０１によって、図２２に示すような同じデータ数の２セグメントに分割して保持され、第２のフレーム分割装置１７０２、第３のフレーム分割装置１７０３、第４のフレーム分割装置１７０４も同様に各々の帯

域信号２，３，４を２セグメントに分割して保持する。

【００１４】そして、再生速度検出装置で得られる目標速度比F から、２セグメントをオーバラップするデータ長である目標オーバラップ値Tbを、次式のように求める。

【００１５】Tb = Tf ・(1 - 1/F)

ここで、後述する位相調整を行う影響による目標速度比F からのずれを補正するための補正パラメータB （初期値は0 ）を考慮して、相関関数算出装置１７１０で、上記第１のフレーム分割装置１７０１の２セグメント間のオーバラップ区間データ長が（Tb + B）の前後mサンプルの範囲で相関関数を演算し、該相関関数が最大となる場合のオーバラップ区間長Tcを求める。その結果、TcがTbからずれることによる目標速度比からの誤差の補正を行うため、先に述べた補正パラメータB の値を以下のように更新する。

【００１６】B ← B + Tb - Tc

図２２は、目標速度比F が2.0 の場合の，目標オーバラップ値Tb （ = Tf/2)の位置関係で、２セグメントを上下に配置した図であり、この場合に2 セグメント間の相関関数を求めた結果が、図２３のようになる。この例では相関が最大値となるような時刻は、４であることがわかる。図２４は、この相関関数の結果をもとに、２セグメント間のオーバラップ長をTcとして表した説明図である。つまり、相関関数によって、前半のセグメントに後半のセグメントの類似度合を求め、その結果、相関の高い位置までずらせると、双方のセグメントの位相が一致することになる。そのときの２セグメント間のオーバーラップ区間長が、Tcということになる。

【００１７】次に、第１のクロスフェード処理装置１７０５で、第１のフレーム分割装置１７０１によって分割され保持された２セグメントの帯域信号を、 Tc 分オーバーラップさせてクロスフェード処理を行なう。同様に、第２のクロスフェード処理装置１７０６、第３のクロスフェード処理装置１７０７、第４のクロスフェード処理装置１７０８でも、それぞれ、第２のフレーム分割装置１７０２、第３のフレーム分割装置１７０３、第４のフレーム分割装置１７０４によって分割され保持された２セグメントの帯域信号を、Tc分オーバーラップさせてクロスフェード処理を行なう。図２５は、このようなクロスフェード処理の一例を示したものである。２セグメントのオーバラップ部分に対して、互いに相補的な重みを付けた加算を行う。（ａ）は、前半のセグメントにフェードアウト処理した信号、（ｂ）は、後半のセグメントにフェードイン処理した信号である。このフェードアウト処理した信号（ａ）と、フェードイン処理した信号（ｂ）とを加算することにより、同図（ｃ）のような波形となる。その後に、帯域合成フィルタ１００９によって、上記のようにしてクロスフェード処理された各帯域信号が帯域合成され、通常の時間軸信号が生成される。

【００１８】以上の処理を、逐次、Ｔｆサンプルずつの全てのフレームに関して、各帯域の信号に行うことによって、１フレーム内のデータだけで完結する高速再生が行えることとなる。

【００１９】しかしながら、上記のような構成による再生装置では、次のような課題が存在する。ここでは、標準的なMPEG1 オーディオの符号化方式を例に取り、分割帯域数を32、各帯域１フレームのデータ数を36、補正パラメータ値Ｂの初期値を０、基準とする相関探索幅mを４として、実際に取り得るオーバラップ値と、相関探索する点数などを、以上に述べた従来例の方法で求め、その結果を以下の表１に示す。ここで、計算式の小数点は切り捨てて表示している。

【００２０】

【表１】

目標速度比 F	1.1	1.3	1.5	1.7	1.9	2.0
目標ｵｰﾊﾞｰﾗｯﾌﾟ値 Tb	3	6	12	14	17	18
ｵｰﾊﾞｰﾗｯﾌﾟ値の範囲	0～7	2～10	8～16	10～18	13～18	14～18
相関探索する点数	8	9	9	9	6	5

【００２１】まず、速度比が1.0 に近い側に関して検討する。目標オーバラップ値が小さいこともあり、オーバラップ値の取り得る値は、かなり小さい値の範囲に留まっている。この場合の問題点として考えられるのは、クロスフェード長が短すぎることである。相関の高い位置を求めてクロスフェード処理を実施するが、クロスフェード区間を挟んだ２セグメント間の遷移期間の長さが短すぎると、セグメント中に含まれる低周波数信号は、クロスフェードによる振幅の連続性改善の効果も少なく、波形の急激な変化をもたらしてしまい、結果として不連続感の強い再生音として聴取される。このクロスフェード区間長および相関探索幅と、音質に関する評価実験は、例えば、「鈴木、三崎：電子情報通信学会音声研究会 SP90-34, 1990.8」などにPCM 音声に対する最適な値を求めている。

【００２２】一方、速度比が2.0 に近い側に関して検討すると、目標オーバラップ値が上限値である18に近い値であり、オーバラップ値の上限が１セグメント長を超えることができず、相関探索点数は十分な数になっていることがわかる。また、速度比2.0 の場合、オーバラップ値を目標値である18より小さい値にとると、次回以降にこれを修正する可能性は全く無いため、目標速度を達成するには、相関による探索は行わずに固定のオーバラップ値を取らざるを得ない。また、相関探索する点数を増

加させるために、探索幅mを大きな値にすると、目標オーバラップ値から小さい側にずれた場合の補正パラメータ値B は正の値であり、次回の相関探索の中心とするオーバラップ値（Tb + B）の値が１セグメント長を超える（（Tb +B）＞; Tf/2），という不合理が生じ、速度比を修正することが困難な状況となる。そのため、探索幅mを小さな値で使用せざるを得なく、相関探索する点数が少ないため、位相の整合性が十分に改善し得ない位置でのクロスフェードを行うことになり、結果として、位相の不整合によりしわがれた声として聴取される。

【００２３】このように、このアルゴリズムでは、相関関数を用いた位相の調整を行うには、不十分な状況で動作させざるを得ないため、十分な性能を出し得ていない。さらに、以上の中間である速度比1.5 近傍の比較的良好と思われる速度の範囲においても、与えられたすべてのフレームに対してクロスフェード処理を実施することになるため、処理による信号劣化が全フレームすべてに生じ、その結果、劣化の度合が大きく感じられることになる。このように、相関関数による位相の整合性を改善する手法は、この例では十分に機能せず、かえって、目標の速度比に収束し難い，という方式上の欠点を有している。また、この例では、高速再生に対する処理を実施するのみで、低速再生に関する機能を何ら提供し得ないものである。

【００２４】本発明は、上記課題に鑑み、フレーム内データで完結する，一定速度比の時間軸圧縮処理または時間軸伸長処理を基本とした簡素な構成によって速度変換処理を行うことができ、高品質な高速または低速の速度変換音声を実現することのできる，音声再生装置を提供することを目的とするものである。

【００２５】

【課題を解決するための手段】この目的を達成するために、請求項１にかかる音声再生装置は、音声復号化手段、選択手段、フレームシーケンステーブル、フレームカウント手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、音声復号化手段は、入力される音声信号をフレーム単位で復号し、選択手段は、与えられる速度比に対応したフレームシーケンスをフレームシーケンステーブルへ出力すると共に、該フレームシーケンスのフレームサイクル数をフレームカウント手段へ出力し、　フレームシーケンステーブルは、選択手段からのフレームシーケンスを記憶し、　フレームカウント手段は、フレームサイクル数に基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、データ伸縮制御手段は、フレームカウント手段のカウント値に対応したフレームシーケンステーブルのフレームシーケンスを参照して、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行うことを特徴とする音声再生装置としたものである。

【００２６】また、請求項２にかかる音声再生装置は、請求項１に記載の音声再生装置において、音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された音声信号を復号することを特徴とする音声再生装置としたものである。

【００２７】また、請求項３にかかる音声再生装置は、請求項１に記載の音声再生装置において、フレームシーケンスは、連続する時間軸圧縮フレームのフレーム数と、連続する時間軸処理無しフレームのフレーム数がいずれも最小となるよう配置されたことを特徴とする音声再生装置としたものである。

【００２８】また、請求項４にかかる音声再生装置は、請求項１に記載の音声再生装置において、フレームシーケンスは、連続する時間軸伸長フレームのフレーム数と、連続する時間軸処理無しフレームのフレーム数がいずれも最小となるよう配置されたことを特徴とする音声再生装置としたものである。

【００２９】また、請求項５にかかる音声再生装置は、音声復号化手段、伸縮頻度制御手段、フレームカウント手段、エネルギー演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された符号化音声信号を復号し、伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、エネルギー演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギーを推定し、フレーム選択手段は、フレームサイクル数Ｎｆのフレーム内でエネルギーの小さいフレームからＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行うことを特徴とする音声再生装置としたものである。

【００３０】また、請求項６にかかる音声再生装置は、音声復号化手段、伸縮頻度制御手段、フレームカウント手段、定常性演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２

符号化方式にて符号化された音声信号を復号し、伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、定常性演算手段は、符号化音声信号のスケールファクタ選択情報をもとにフレームサイクル数Ｎｆ分の符号化音声信号の定常性を推定し、　フレーム選択手段は、フレームサイクル数Ｎｆのフレーム内での定常性の高いフレームからＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行うことを特徴とする音声再生装置としたものである。

【００３１】また、請求項７にかかる音声再生装置は、音声復号化手段、伸縮頻度制御手段、フレームカウント手段、エネルギー変化度合演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された音声信号を復号し、伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、エネルギー変化度合演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギー変化度合を推定し、フレーム選択手段は、フレームサイクル数Ｎｆのフレーム内でエネルギー変化度合に基づき継時マスキング効果による処理劣化が少ないフレームからＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行うことを特徴とする音声再生装置としたものである。

【００３２】また、請求項８にかかる音声再生装置は、音声復号化手段、伸縮頻度制御手段、フレームカウント手段、演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された符号化音声信号を復号し、伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、演算手段は、エネルギー演算手段、定常性演算手段、エネルギー変化度合演算手段のいずれか２つ以上を備え、エネルギー演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギーを推定し、定常性演算手段は、符号化音声信号のスケールファクタ選択情報をもとにフレームサイクル数Ｎｆ分の符号化音声信号の定常性を推定し、エネルギー変化度合演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギー変化度合を推定し、フレーム選択手段は、演算手段の出力をもとにＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行うことを特徴とする音声再生装置としたものである。

【００３３】また、請求項９にかかる音声再生装置は、請求項１～８のいずれかに記載の音声再生装置において、データ伸縮手段は、クロスフェード手段を備え、クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを重み付け加算することを特徴とする音声再生装置としたものである。

【００３４】また、請求項１０にかかる音声再生装置は、請求項１～８のいずれかに記載の音声再生装置において、データ伸縮手段は、相関演算手段、クロスフェード手段を備え、相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、セグメント間の相関値を演算し、相関値が高くなる位置で重み付け加算するためのシフト量を決定し、クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算することを特徴とする音声再生装置としたものである。

【００３５】また、請求項１１にかかる音声再生装置は、請求項１～８のいずれかに記載の音声再生装置において、音声復号化手段は、符号化音声信号を帯域毎に復号し、データ伸縮手段は、相関演算手段、帯域毎のクロスフェード手段を備え、相関演算手段は、音声復号化手

段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、ピッチ周波数を包含する帯域においてセグメント間の相関値を演算し、相関値が高くなる位置で重み付け加算するためのシフト量を決定し、各クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算することを特徴とする音声再生装置としたものである。

【００３６】また、請求項１２にかかる音声再生装置は、請求項１～８のいずれかに記載の音声再生装置において、音声復号化手段は、符号化音声信号を帯域毎に復号し、データ伸縮手段は、相関演算手段、帯域毎のクロスフェード手段を備え、相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、平均エネルギーが最大となる帯域においてセグメント間の相関値を演算し、相関値が高くなる位置で重み付け加算するためのシフト量を決定し、各クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算することを特徴とする音声再生装置としたものである。

【００３７】また、請求項１３にかかる音声再生装置は、請求項１～８のいずれかに記載の音声再生装置において、音声復号化手段は、符号化音声信号を帯域毎に復号し、データ伸縮手段は、相関演算手段、帯域毎のクロスフェード手段を備え、相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、各帯域においてセグメント間の相関値を演算し、相関値が最大の帯域において相関値が高くなる位置で重み付け加算するためのシフト量を決定し、各クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算することを特徴とする音声再生装置としたものである。

【００３８】

【００３９】

【００４０】

【００４１】

【００４２】

【００４３】

【００４４】

【００４５】

【発明の実施の形態】（実施の形態１）以下、本発明の第１の実施の形態について、図面を参照しながら説明する。図１は本発明の第１の実施の形態における音声再生装置のブロック図を示すものである。図１において、１０１はフレーム逆パッキング手段、１０２は逆量子化手段、１０３はデータ伸縮手段、１０４はサブバンド合成フィルタ手段、１０５は選択手段、１０６はフレームカウント手段、１０７はデータ伸縮制御手段、１０８はフレームシーケンステーブルである。以下に、その動作について説明する。

【００４６】本実施の形態は、MPEG１オーディオのビットストリームをデコードする際の中間データに対して速度変換処理を施す音声再生装置の例を示すものである。MPEG１オーディオのビットストリームは、ヘッダ，ビット割当て情報，スケールファクタに関する情報，サンプルデータ情報などから成り立っている。

【００４７】図１において、入力されたMPEG１オーディオのビットストリームは、フレーム逆パッキング手段１０１によって、当該ビットストリームからヘッダ，ビット割当て情報，スケールファクタに関する情報，サンプルデータ情報などの個々の情報に分離される。逆量子化手段１０２では、当該逆パッキングにて得られた、各帯域（MPEG１オーディオでは３２のサブバンド（帯域））毎のビット割当て情報や、スケールファクタに関連する情報をもとにして、各帯域毎に逆量子化したデータを得る。

【００４８】データ伸縮手段１０３は、後述するデータ伸縮制御手段１０７からの制御によって、時間軸圧縮／伸長を施すフレームに該当する時は、逆量子化手段１０２の出力を一定比率で時間軸圧縮／伸長し、圧縮／伸長することなくスルーで出力するフレームに該当する場合には、逆量子化手段１０２の出力をそのままサブバンド合成フィルタ出力手段１０４へ出力する。サブバンド合成フィルタ手段１０４では、入力された各サブバンド（MPEG１オーディオでは３２帯域）のデータが帯域合成され、当該合成により得られたオーディオ信号を出力する。

【００４９】図２に、データ伸縮手段１０３の内部構成図を示す。同図において、２００１は最も低いサブバンドに対応する逆量子化手段１０２の出力Q0に対して処理を施すデータ伸縮ユニット０であり、以降、２００２，・・・，２０３２の各々は、各サブバンドの低い帯域側からの逆量子化手段１０２の出力Q1，・・・,Q31に対して処理を施す、データ伸縮ユニット１、・・・、データ伸縮ユニット３１である。データ伸縮ユニットの内部の構成は、図２に示すように、バッファメモリ２０１，クロスフェード手段２０２，データ選択手段２０３で構成される。図では、データ伸縮ユニット１～データ伸縮ユニット３１については内部構成が記載されていないが、データ伸縮ユニット０と同一であるので、図では省略して記載している。

【００５０】以下では、最も低い周波数帯域に相当する逆量子化手段１０２の出力データQ0に対して処理を施すデータ伸縮ユニット０の動作を示す。逆量子化手段１０２の出力Q0は、一旦バッファメモリ２０１に１フレーム

分（所定時間長分）のデータだけ蓄積される。ここで、各サブバンドにおける１フレームのデータ数を、Nsとする。データ伸縮制御手段１０７からの制御信号により、スルーで出力するフレームに該当する場合には、データ選択手段２０３は、バッファメモリ２０１へ書き込まれているNs個のデータを、そのままサブバンド合成フィルタ手段１０４へ出力する。一方、データ伸縮制御手段１０７からの制御信号により、時間軸圧縮／伸長を施すフレームに該当する時は、クロスフェード手段２０２にて、バッファメモリ２０１内のNs個のデータを用いて、所定の圧伸比Srで時間軸圧縮／伸長を行なう。

【００５１】クロスフェード手段によるクロスフェード処理、すなわち時間軸圧縮／伸長の方法を、図３を用いて説明する。図３は、時間軸圧縮／伸長を実施することで、フレームのデータ長が変化する様子を示した模式図の一例である。図３（ａ）は通常のフレームを示すものであり、ここでは、１フレームのデータ数Nsを、同数のデータ数（同一時間長）のセグメントであるＳＥＧ１，ＳＥＧ２の２つに分割した例である。これらのセグメントを基にして、図３に示すような重み付け加算、すなわち、クロスフェード処理を行うことによって、前後の不連続無く、データ数を減少／増加させることができる。例えば、圧縮する場合は図３（ｂ）のように行い、伸長する場合は同図（ｃ）のようにクロスフェード処理を行う。また、クロスフェード処理を施すことなくデータ伸縮手段１０３にてスルーで出力するフレームである場合には、図３（ａ）に示すフレーム信号がそのままの状態でサブバンド合成フィルタ手段１０４へ出力される。なお同図において、（ｂ）は圧伸比（＝１／速度比）１／２で時間軸圧縮されたフレームの例、（ｃ）は圧伸比３／２で時間軸伸長されたフレームの例である。なお、圧伸比は、圧伸比＝１／速度比＝クロスフェード手段からの出力データ数／クロスフェード手段への入力データ数で定義するものとする。

【００５２】図３（ｂ）のような圧縮処理を全フレームに対して行うことにより、一定の速度比２．０の再生音を得ることができる。また、同図（ｃ）のような伸長処理を全フレームに対して行うことにより、一定の速度比２／３の再生音を得ることができる。このような速度変換処理を行う場合には、データ伸縮制御手段１０７からデータ伸縮手段１０３へ、圧縮／伸長／スルーを示す制御信号を送り、この制御信号を基に、各データ伸縮ユニットを制御することにより、該速度変換処理を実現できる。例えば、上述したような速度比２．０を実現するには、入力された速度比情報（＝２．０）を基に、「速度比＝２．０の圧縮」を示す制御信号をデータ伸縮手段１０３へ出力する。クロスフェード手段２０２は、当該制御信号を受けとって、全フレームに対して図３（ｂ）に示すクロスフェード処理を行い、データ選択手段２０３は、クロスフェード手段２０２の出力を選択して、サブバンド合成フィルタ手段１０４へ出力する。また、速度比２／３（＝０．６６）を実現するには、入力される速度比情報（＝２／３）を基に、「速度比＝２／３の伸長」を示す制御信号を、データ伸縮手段１０３へ出力する。クロスフェード手段２０２は、当該制御信号を受けとって、全フレームに対して図３（ｃ）に示すクロスフェード処理を行い、データ選択手段２０３は、クロスフェード手段２０２の出力を選択して、サブバンド合成フィルタ手段１０４へ出力する。

【００５３】上述した以外の速度比の音声を実現するためには、全フレームではなく特定のフレームに対してのみ、図３（ｂ）（ｃ）のような時間軸圧縮／伸長を行うシーケンスで処理を繰り返せば、最終的には個々のフレームの速度比とは異なった，所望の再生速度を得ることが可能となる。図４を用いてこの例を説明する。

【００５４】図４は、その一例として、速度比が1.5，1.2，1.1，0.9，0.8，0.7 の場合の時間軸圧縮／伸長処理を説明するための処理シーケンス図である。同図において、（ａ）はスルー（時間軸圧縮／伸長処理なし）で出力するフレーム、（ｂ）は時間軸圧縮処理を施すフレーム、（ｃ）は時間軸伸長処理を施すフレームを示している。また（表２）に、図４の処理速度例における、入力セグメント数，出力セグメント数，圧縮／伸長するセグメント数，繰り返しを行うフレームサイクルを示す。図４における各フレームは、図３（ａ）にて説明したように、同一データ数（同一時間長）の２つのセグメントから構成されており、よって、各速度比における入力・出力セグメント数，圧縮／伸長セグメント数は、（表２）の通りとなる。例えば、速度比1.5 の場合を例にとると、入力セグメント数は、図４（イ）の通り第１フレーム～第３フレームが入力されるので、３フレーム×２セグメント＝６セグメントである。このとき、第２フレーム、第３フレームについては、時間軸圧縮処理が施されて各フレームのセグメント数が２→１になるので、圧縮セグメント数は２となり、この結果、出力セグメント数は、６セグメント－２セグメント＝４セグメントとなる。速度比は（入力セグメント数／出力セグメント数）で与えられる。また（表３）に、図４に対応する、フレームシーケンステーブル１０８に与えるべきデータ例を示す。この例においては、テーブルには、速度比と、フレームカウント手段１０６でカウントするフレームサイクルと、フレームに対する圧縮／伸長／スルーの処理状態のシーケンス（フレームシーケンス）とが、記録されている。

【００５５】なお、（表３）において、「ａ」はスルー、「ｂ」は圧縮、「ｃ」は伸長、を施すシーケンスを意味している。

【００５６】

【表２】

速度比	0.7	0.8	0.9	1.1	1.2	1.5
入力ｾｸﾞﾒﾝﾄ数	14	4	18	22	6	6
出力ｾｸﾞﾒﾝﾄ数	20	5	20	20	5	4
圧縮/伸長ｾｸﾞﾒﾝﾄ数	6	1	2	-2	-1	-2
ﾌﾚｰﾑｻｲｸﾙ	7	2	9	11	3	3

【００５７】

【表３】

速度比	ﾌﾚｰﾑｻｲｸﾙ	ﾌﾚｰﾑｼｰｹﾝｽ
0.7	7	a,c,c,c,c,c,c
0.8	2	a,c
0.9	9	a,a,c,a,a,a,c,a,a
1.1	11	a,a,b,a,a,a,a,a,b,a,a
1.2	3	a,b,a
1.5	3	a,b,b

a：スルー
b：圧縮
c：伸長

【００５８】まず、所望の速度比情報が選択手段１０５へ入力される。本例の場合においては、速度比＝1.1，速度比＝0.7 などの情報である。この速度比情報が入力されると、選択手段１０５は、フレームカウント手段１０６へフレームサイクルを、フレームシーケンステーブル１０８へはフレームシーケンスを送出する。この際に送出されるフレームサイクル，フレームシーケンスは、（表３）に示されるような値である。

【００５９】以下、再生時間を短くする（速度比＞1.0；時間軸圧縮処理）例を、速度比1.1 の場合を例にとって説明する。

【００６０】速度比情報1.1 が選択手段１０５へ入力されると、選択手段１０５はフレームカウント手段１０６へフレームサイクル「１１」を、またフレームシーケンステーブル１０８へフレームシーケンス「ａ，ａ，ｂ，ａ，ａ，ａ，ａ，ａ，ｂ，ａ，ａ」を送出する。このフレームシーケンスは、フレームシーケンステーブル１０８に書き込まれる。フレームカウント手段１０６は、選択手段１０５からフレームサイクル「１１」を受け取ったタイミング以降に、フレーム逆パッキング手段１０１から出力されフレームカウント手段１０６へ入力されたフレームをカウントし、フレームカウント値を出力する。この際、フレームカウント手段１０６のカウント値は、１→２→・・・→１０→１１→１→・・・と、１１サイクルでカウントされるものとする。

【００６１】データ伸縮制御手段１０７は、当該カウント値をもとに、まず、カウント値「１」が入力されたときはフレームシーケンステーブル１０８からフレームシーケンス１番目のシーケンス「ａ」を読み込み、データ伸縮手段１０３へ「スルー」を指示する制御信号を出力する。データ伸縮手段１０３において、当該手段内部の各データ選択手段は、この「スルー」を意味する制御信号を基に、逆量子化手段１０２から出力されたQ0，Q1，・・・，Q31 なるデータを、スルー（クロスフェード処理なし）で出力する（C0，C1，・・・，C31 ）。サブバンド合成フィルタ手段１０４では、当該３２帯域のC0，C1，・・・，C31 を基にして帯域合成し、当該フレームのオーディオ出力として出力される。

【００６２】次に、フレームカウント手段１０６からカウント値「２」が出力されると、データ伸縮制御手段１０７は、フレームシーケンステーブル１０８からフレームシーケンス２番目のシーケンス「ａ」を読み込み、データ伸縮手段１０３へ「スルー」を指示する制御信号を出力する。これ以降の処理は、上述したカウント値「１」の場合と同じである。なお、図４、（表２）からも明らかなように、カウント値「４」「５」「６」「７」「８」「１０」「１１」の場合にフレームシーケンステーブル１０８から読み込まれるシーケンスは「ａ」であり、この際の処理は上述したカウント値「１」の場合と同一なので、説明を省略する。

【００６３】図４および（表３）より、フレームカウント値が「３」，「９」の場合には、フレームシーケンステーブル１０８からは、フレームシーケンスとして「ｂ」が読み込まれ、時間軸圧縮処理が施される。これについて、以下に説明する。

【００６４】フレームカウント手段１０６からカウント値「３」，「９」が出力された場合には、データ伸縮制御手段１０７は、フレームシーケンステーブル１０８からフレームシーケンス「ｂ」を読み込み、これによりデータ伸縮手段１０３へ「圧縮」を指示する制御信号を出力する。データ伸縮手段１０３において、当該手段内部の各データ選択手段は、この「圧縮」を意味する制御信号を基に、データ伸縮ユニット０～データ伸縮ユニット３１内の各クロスフェード手段にて、図３（ｂ）を用いて上記説明した時間軸圧縮処理を行い、この圧縮処理が施された信号がデータ選択手段２０３にて選択されて、出力される（C0，C1，・・・，C31 ）。サブバンド合成フィルタ手段１０４では、当該３２帯域のC0，C1，・・・，C31を基にして帯域合成し、当該フレームのオーディオ出力として出力される。

【００６５】上述のような処理にて各フレームに対して、スルー、時間軸伸長の処理が成され、フレームシーケンス「１１」で１サイクルの処理が行われる。この１サイクル処理が終わると、その後入力されてくる各フレームに対して、上述したシーケンスと同一のシーケンスで処理が継続される。

【００６６】次に、再生速度を遅くする（速度比＜1.0;時間軸伸長処理）例を、速度比0.7の場合を例にとって説明する。速度比情報0.7 が選択手段１０５へ入力されると、選択手段１０５はフレームカウント手段１０６へフレームサイクル「7」を、またフレームシーケンステーブル１０８へフレームシーケンス「ａ，ｃ，ｃ，ｃ，ｃ，ｃ，ｃ」を送出する。このフレームシーケンスは、フレームシーケンステーブル１０８に書き込まれる。フレームカウント手段１０６は、選択手段１０６からフレームサイクル「7」を受け取ったタイミング以降に、フレーム逆パッキング手段１０１から出力されフレームカウント手段１０６へ入力されたフレームをカウントし、フレームカウント値を出力する。この際、フレームカウント手段１０６のカウント値は、１→２→・・・→６→７→１→・・・と、７サイクルでカウントされるものとする。

【００６７】データ伸縮制御手段１０７は、当該カウント値をもとに、まず、カウント値「１」が入力されたときは、フレームシーケンステーブル１０８からフレームシーケンス１番目のシーケンス「ａ」を読み込み、データ伸縮手段１０３へ「スルー」を指示する制御信号を出力する。データ伸縮手段１０３において、当該手段内部の各データ選択手段は、この「スルー」を意味する制御信号を基に、逆量子化手段１０２から出力されたQ0，Q1，・・・，Q31 なるデータを、スルー（クロスフェード処理なし）で出力する（C0，C1，・・・，C31 ）。サブバンド合成フィルタ手段１０４では、当該３２帯域のC0，C1，・・・，C31 を基にして帯域合成し、当該フレームのオーディオ出力として出力される。

【００６８】次に、フレームカウント手段１０６からカウント値「２」が出力されると、データ伸縮制御手段１０７は、フレームシーケンステーブル１０８からフレームシーケンスとして「ｃ」が読み込まれ、時間軸伸長処理が施される。これについて、以下に説明する。

【００６９】フレームカウント手段１０６からカウント値「２」が出力された場合には、データ伸縮制御手段１０７は、フレームシーケンステーブル１０８からフレームシーケンス「ｃ」を読み込み、これによりデータ伸縮手段１０３へ「伸長」を指示する制御信号を出力する。データ伸縮手段１０３において、当該手段内部の各データ選択手段は、この「伸長」を意味する制御信号を基に、データ伸縮ユニット０～データ伸縮ユニット３１内の各クロスフェード手段にて、図３（ｃ）を用いて上記説明した時間軸伸長処理を行い、この伸長処理が施された信号がデータ選択手段２０３にて選択されて、出力される（C0，C1，・・・，C31 ）。サブバンド合成フィルタ手段１０４では、当該３２帯域のC0，C1，・・・，C31 を基にして帯域合成し、当該フレームのオーディオ出力として出力される。

【００７０】次に、フレームカウント手段１０６からカウント値「３」が出力されるが、図４および（表３）からも明らかなように、カウント値「３」「４」「５」「６」「７」の場合にフレームシーケンステーブル１０８から読み込まれるシーケンスは、第２フレームと同様に「ｃ」であり、この際の処理は上述したカウント値「２」の場合と同一なので、説明を省略する。

【００７１】上述のような処理にて、各フレームに対して、スルー、時間軸伸長の処理が成され、フレームシーケンス「7」で１サイクルの処理が行われる。この１サイクル処理が終わると、その後入力されてくる各フレームに対して、上述したシーケンスと同一のシーケンスで処理が継続される。

【００７２】以上の説明より明らかなように、フレームサイクルで所望の速度比のデータ数（セグメント数）になるように、時間軸圧縮／伸長するフレームを偏りがあまりないように挿入することにより、特定のフレームサイクル内で所望の速度比を得ることが可能となる。また図４，（表２）（表３）の例とは異なる速度比の場合でも、速度比に合うように、時間軸圧縮／伸長するフレームを偏りがあまりないように挿入するシーケンステーブルを用いてフレームサイクルを繰り返すことにより、所望の速度比の音声を得ることが可能である。また、図４，（表２）（表３）の例とは異なる順番であるシーケンスパターンの場合でも、（表２）に示すような圧縮／伸長セグメント数が守られておれば、所望の速度比が得られる。

【００７３】このように、一定値（本実施形態では図３のように圧縮比は１／２、伸長比は３／２）の時間軸圧縮／伸長を行うフレームを所定の順番で実施するように制御すれば、所望の速度比の音声を得ることが可能となる。

【００７４】なお、以上の説明においては、図３に示したように、基準とする時間軸圧縮比の値を１／２，時間軸伸長比の値を３／２で実現した例で説明したが、その他の時間軸圧縮比／伸長比をもとにシーケンステーブルを構成することも、同様に実施可能である。

【００７５】（実施の形態２）以下、本発明の第２の実施の形態について、図面を参照しながら説明する。第２の実施の形態における音声再生装置の構成図は、上述した第１の実施の形態の構成図（図１）と基本的に同様の構成であり、MPEG１オーディオストリームを入力する例である。フレーム逆パッキング手段１０１、逆量子化手段１０２、サブバンド合成フィルタ手段１０４、選択手段１０５、フレームカウント手段１０６、フレームシー

ケンステーブル１０８、データ伸縮制御手段１０７、は第１の実施の形態と同様の動作をするものである。本第２の実施の形態が第１の実施の形態と異なっている点は、データ伸縮手段１０３の内部の構成および動作にある。

【００７６】本第２の実施の形態におけるデータ伸縮手段の構成図を、図５に示す。

【００７７】同図において、２００１は最も低いサブバンドに対応する逆量子化手段１０２の出力Q0に対して処理を施すデータ伸縮ユニット０であり、以降、２００２，・・・，２０３２の各々は、各サブバンドの低い帯域側からの逆量子化手段１０２の出力Q1，・・・,Q31に対して処理を施す、データ伸縮ユニット１、・・・、データ伸縮ユニット３１である。各データ伸縮ユニットの内部の構成は、図５に示すように、バッファメモリ２０１，クロスフェード手段２０２，データ選択手段２０３で構成される。図ではデータ伸縮ユニット１～データ伸縮ユニット３１については内部構成が記載されていないが、データ伸縮ユニット０と同一であるので、図では省略している。本実施形態の構成は、さらに、図５に示すように、相関演算手段３０１、位相制御記憶手段３０２を付加した構成となっている。

【００７８】以下に、相関演算手段３０１と、位相制御記憶手段３０２の動作を中心に説明を行う。第１の実施形態においては、時間軸波形のクロスフェードは一意に定位置で重み付け加算されている。この場合、波形の振幅に関しては不連続無く接続されるが、位相に関しては考慮されていない。そこで、本実施形態においては、位相の整合性が高い位置を相関関数を用いて求め、その位置にシフトしてから重み付け加算を行うようなクロスフェード処理を行うようにする。図６に、このような重み付け加算を行ったクロスフェード処理（圧縮）の例を示す。図６（ａ）は、図３（ａ）に相当する、クロスフェード処理を施す前の元のフレームを示しており、同一データ数のセグメント１と、セグメント２とから成っている。図６（ｂ）は、セグメント１と、セグメント２が、相関を考慮したシフトが成されることなく重み付け加算されており、これを図３（ｂ）の圧縮フレームと同一の基準形と考える。図６（ｃ）は、相関の高い位置が基本形の場合に比べて右に存在した場合のクロスフェード処理後のフレームであり、クロスフェード区間は、同図（ｂ）の基準形に比べて短くなるとともに、データ量は、同図（ｂ）に比べて増加する。逆に図６（ｄ）は、相関の高い位置が左に存在した場合のクロスフェード処理後のフレームであり、クロスフェード区間は、同図（ｂ）の基準形に比べて短くなるとともに、データ量は同図（ｂ）に比べて減少する。

【００７９】上述の如き、位相の整合性を改善する目的のために、相関関数を用いたクロスフェード処理を行う速度変換装置については、本願出願人により種々の提案が成されており、例えば、本願出願人の先願たる特開平４－１０４２００号公報（特許登録２５３２７３１号）などに示される通りである。本実施形態では、このような相関関数を用いたクロスフェード手法を用いるが、この際図５において、最も低域のデータであるQ0には、音声のピッチ周波数が存在する範囲が含まれると考えられるので、このピッチ周波数に相当する成分に関して位相の整合性を改善するために、Q0に相当する帯域データのみを用いて、相関演算手段３０１、位相制御記憶手段３０２により相関演算を行う。相関演算を行うデータは、バッファメモリ２０１に存在しているが、相関演算の範囲は、上記した特開平４－１０４２００号公報などに示されるように、与えられたフレームシーケンスの値が、圧縮フレームか伸長フレームのいずれであるかと、前回求めた相関シフト量とによって決定される。

【００８０】図６（ｃ）（ｄ）からもわかるように、相関の高い位置にシフトした場合には、本来目標としているデータ数（図６（ｂ））に比較して過不足を生じることになる。その過不足の値は、相関の高い位置にシフト（相関シフト量をrkとする）したデータ量から求めることができ、これを次回生じる時間軸圧縮／伸長処理の際に補うことにする。そのためには、データの過不足に相当する相関シフト量rkを、一旦、位相制御記憶手段３０２に記憶する必要がある。このシフト量rkは、次回のクロスフェード処理を行う際の、加算する先頭データの位置（ポインタ）を調整することにより、補正できることになる。

【００８１】このようなシフト量rkの補正を行う様子を、図７に模式的に示す。以前の圧縮フレームにおいて、基準形（図７（ａ））のようにシフトが生じなかった場合、ポインタP2の位置のシフトは無く、図のような位置関係で相関の高い位置を探索するので、今回の基準形でも、セグメント１と、セグメント２は、シフト無くクロスフェード処理される。以前の圧縮フレームにおいて、正方向にシフト（rk>0）した位置で重み付け加算が行われた場合（図７（ｂ））、以前に余分にデータを出力しているので、今回のポインタ位置はP2が正方向にシフトした位置となり、今回の基準形では、セグメント１内の後ろ部分と、セグメント２内の前部分とが使用されないことになり、よってこの際の基準形は、図７（ｂ）の如くになる。また、以前の圧縮フレームにおいて、負方向にシフト（rk<0）した位置で重み付け加算が行われた場合（図７（ｃ））、以前にデータを不足させているので、今回のポインタ位置は、P2が負方向にシフトした位置となり、今回の基準形では、セグメント１内の後ろ部分は複数回（この場合２回）使用されることになり、よってこの際の基準形は、図７（ｃ）の如くになる。いずれの場合でも、図７に示すような処理を施すことによって、今回の伸縮フレームにおける基準形の圧縮が行われる時には、以前のフレームのデータ量の、目

標とするデータ量に対する誤差は吸収されていることになり、よって誤差の累積は無いことになる。上述の例では、圧縮フレームに関して説明を行ったが、伸長フレームに関しても同様の考え方で実現できることは言うまでもない。このように、以前の圧縮／伸長フレームのシフト量を考慮して、ポインタ位置をシフトした位置を基準として、相関関数で相関の高い位置を求めることになる。

【００８２】以上のように求められた相関シフト量rkは、他のサブバンドにおいても同様に適用してクロスフェード処理が行われ、Q0に対するクロスフェード処理と同様の処理がQ1～Q31 に対しても行われる。これにより、各サブバンドにて、同一のシフト量rkにてクロスフェード処理が施されたのち、C0～C31 の出力信号が合成されることになる。

【００８３】以上のように、本実施の形態２の構成によれば、相関演算手段３０１によって位相の整合性の高い位置で重み付け加算を行うクロスフェード処理を行うことで、データ伸縮手段１０３の出力信号の振幅・位相の両方が、前後のフレームに対して不連続無く接続されるため、音質の向上を達成することができる。

【００８４】なお、上記実施の形態２では、最低域のサブバンドの逆量子化出力データQ0に対して相関関数を求めており、音声に対する基本周波数を元に、位相の整合性を改善することに主眼をおいているが、MPEG符号化などの音声（スピーチ）信号以外の音源の場合には、必ずしも、最低域のサブバンドについて相関関数を求めることが良い結果をもたらすとは限らない。そのため、各サブバンドの逆量子化手段の出力データのすべて（第１、第２の実施形態の例でいうなら、Q0～Q31 ）に対して相関の高い位置を求め、その各サブバンドの最大相関値の中で最も大きいサブバンドの相関値を元に、重み付け加算するシフト量を決定することにより、周期性の高い帯域を中心とした位相の整合性を改善させることが可能となる。また、各サブバンドの平均エネルギーを求め、その最も平均エネルギーの大きいサブバンドに対して相関の高い位置を求めることによっても、同様の改善を達成することができる。

【００８５】さらに、本実施の形態２の説明で述べたような１つの速度比に対して１つのフレームシーケンスを用いる１対１対応でなく、例えば図８に示すように、１つの速度比に対して伸縮フレームの発生位置が異なる複数のフレームシーケンステーブルを用意しておき（図８の例は速度比が1.1 の場合）、伸縮フレームにおける相関値の平均を、各フレームシーケンステーブル毎に予め求めて、最も相関値の平均が高いシーケンステーブルを参照して伸縮処理を実施するようにして、伸縮フレームを発生させる位置を、より最適な位置のもので行うことにより、位相の整合性の改善度を高めることが可能となる。さらに、先に述べた各帯域における相関値の中で最大相関値を採用する方法と組み合わせれば、一層よい改善を発揮することができる。

【００８６】（実施の形態３）以下、本発明の第３の実施の形態について、図面を参照しながら説明する。図９は本発明の第３の実施の形態による音声再生装置のブロック図を示すものである。図９において、３００１はフレーム復号化手段、３００２はデータ伸縮手段、３００３は伸縮頻度制御手段、３００４はエネルギー演算手段、３００５はフレーム選択手段、３００６はデータ伸縮制御手段である。以下に、その動作について説明する。

【００８７】本実施の形態３は、フレーム単位で復号化処理を行う音声に対して速度変換処理を施す音声再生装置の一例を示すものである。

【００８８】図９において、最初に、伸縮頻度制御手段３００３は、与えられた速度比の情報をもとに、速度変換処理の一連の処理の１周期に相当するフレームサイクル数Nfと、そのフレームサイクル数内で伸縮処理を行うフレーム数Nsとを出力する。そして、エネルギー演算手段３００４では、伸縮頻度制御手段で決定されたフレームサイクル数分の音声のエネルギーを求める。次に、フレーム選択手段３００５は、先に求められたNf個のエネルギーの値を参考に、音声が存在しない無音状態のフレームはエネルギーが小さく、そのフレームを伸縮処理しても劣化は検知され難いと仮定し、速度変換処理のために伸縮すべきフレームとして、エネルギーの小さいフレームから優先的に所定数Ns個の選択を行う。そして、データ伸縮制御手段３００６は、当該フレームが伸縮すべきフレームとして選択されたフレームかどうかを判断し、データ伸縮手段３００２が、伸縮処理をすべきかどうかを制御する。その結果、入力された符号化データは、フレーム復号化手段３００１で１フレーム単位で復号化され、データ伸縮制御手段によって伸縮すべきと判断されたフレームについて、波形の伸縮を行い、それ以外のフレームについては、そのまま出力を行う。このように、あらかじめエネルギー演算手段で求めた音声のエネルギーを用いて、フレーム選択手段で、フレームサイクル内で伸縮すべき最適なフレームを求めておくことにより、速度変換処理音声として、波形の伸縮による処理劣化が検知され難くするようにする，ことが可能となる。

【００８９】なお、本実施の形態３では、各エネルギーの値を参考に、音声が存在しない無音状態のフレームはエネルギーが小さいと仮定し、伸縮すべきフレームとして、エネルギーの小さいフレームから優先的に所定数のフレームを選択するようにしているが、各フレームにおける平均振幅の値を用いる場合にも、有効であると考えられる。

【００９０】（実施の形態４）以下、本発明の第４の実施の形態について、図面を参照しながら説明する。図１

０は本発明の第４の実施の形態による音声再生装置のブロック図を示すものである。図１０において、３００１はフレーム復号化手段、３００２はデータ伸縮手段、３００３は伸縮頻度制御手段、４００４は音声らしさ演算手段、４００５はフレーム選択手段、３００６はデータ伸縮制御手段である。以下に、その動作について説明する。

【００９１】本実施の形態４は、フレーム単位で復号化処理を行う音声に対して速度変換処理を施す音声再生装置の一例を示すものである。図１０において、フレーム復号化手段３００１、データ伸縮手段３００２、伸縮頻度制御手段３００３、データ伸縮制御手段３００６は、実施の形態３と同様の動作を行うものである。本実施の形態4 では、伸縮処理を行うべきフレームの選択を行うフレーム選択手段の働きを、中心に説明を行う。

【００９２】ここでは音声らしさという尺度をもとに、選択すべきフレームの判定を行う。ここで、音声らしさに関して説明を行う。実環境などにおける，通信や放送などでの音声信号においては、全くの無音状態あるいはそれに近い状態という状況は、ほとんどありえない。必ず背景騒音や目的としない音が混入し、目的とする音声信号に重畳する形で含まれている。つまり、より厳密に人間の音声を含むフレームを選択するには、エネルギーの大小だけではなく、含まれるフレームの性質を別の観点で分析する必要がある。そこで、該当するフレームにどのくらいの確からしさで音声信号が含まれているかを推定する尺度として、「音声らしさ」の定義を示す。中藤らによる，「ファジー推論による音声／雑音判別手法の検討」（1993年電子情報通信学会春季大会，A-223 ）による手法で、母音・無声摩擦音の発生頻度をファジー推論することにより、会話の音声らしさを求めて、これと予め求めてある閾値との比較によって、入力信号が、音声／雑音のいずれであるかの2 者択一の判定を行っている。この音声らしさは、特定の時間内に音声が含まれる可能性を示す尺度として用いれば、雑音と音声の混入している音声でも、最も音声が含まれないと予想されるフレームを推定することができる。また、音声らしさの度合を数値化していることにより、複数フレームの音声らしさの大小に基づく相対比較判定に利用することができる。

【００９３】人間が自然に発声する音声を速度別に分析すると、該人間が自然に発生する音声は、言語情報を担う音声区間以外の発声器官が休止しているポーズ区間長を伸縮させている度合が大きいことが判っている（参考文献2 ）参照）。従って、自然な音声速度変換処理を実施するためには、ポーズ区間であるところの非音声区間を伸縮する方が好ましい。

【００９４】音声らしさ演算手段４００４では、伸縮頻度制御手段で決定されたフレームサイクル数分の音声らしさを求める。次に、フレーム選択手段４００５は、先に求められたNf個の音声らしさの値を参考に、音声らしさが小さいフレームは音声情報が少なく、そのフレームを伸縮処理しても劣化は検知され難いと仮定し、速度変換処理のために伸縮すべきフレームとして、音声らしさの小さいフレームから優先的に所定数Ns個の選択を行う。そして、データ伸縮制御手段３００６は、伸縮すべきフレームとして選択されたフレームかどうかを判断し、データ伸縮手段３００２が伸縮処理をすべきかどうかを制御する。その結果、入力された符号化データはフレーム復号化手段３００１で１フレーム単位で復号化され、データ伸縮制御手段によって伸縮すべきと判断されたフレームについて波形の伸縮を行い、それ以外のフレームについては、そのまま出力を行う。このように、あらかじめ、音声らしさ演算手段で求めた音声のエネルギーを用いて、フレーム選択手段でフレームサイクル内で伸縮すべき最適なフレームを求めておくことにより、速度変換処理音声として、波形の伸縮による処理劣化が検知され難いものとすることが可能となる。

【００９５】（実施の形態５）以下、本発明の第５の実施の形態について、図面を参照しながら説明する。図１１は本発明の第５の実施の形態による音声再生装置のブロック図を示すものである。図１１において、３００１はフレーム復号化手段、３００２はデータ伸縮手段、３００３は伸縮頻度制御手段、５００４は定常性演算手段、５００５はフレーム選択手段、３００６はデータ伸縮制御手段である。以下に、その動作について説明する。

【００９６】本実施の形態５は、フレーム単位で復号化処理を行う音声に対して速度変換処理を施す音声再生装置の一例を示すものである。図１１において、フレーム復号化手段３００１、データ伸縮手段３００２、伸縮頻度制御手段３００３、データ伸縮制御手段３００６は、実施の形態３と同様の動作を行うものである。本実施の形態５では、伸縮処理を行うべきフレームの選択を行うフレーム選択手段の働きを中心に説明を行う。

【００９７】本実施の形態５では、音声波形の定常性に着目する。ここでは、フレーム内における正規化自己相関関数を求め、その値の大きいものほど定常性が高いと考える。これは、時間伸縮処理は時間軸波形の類似区間をもとに波形の挿入・間引き操作を行う場合、相関の高いフレームでは波形の重み付け加算による伸縮処理を行うため、処理劣化が検知され難い定常性の高いフレームを選択して、伸縮処理を行うことにする。逆に、音声の子音の始端部分などの非定常な過渡的な部分では、重み付け加算による劣化が顕著となる。

【００９８】定常性演算手段５００４では、伸縮頻度制御手段３００３で決定されたフレームサイクル数分の定常性を、予め求める。次に、フレーム選択手段５００５は、先に求められたNf個の定常性の値を参考に、定常性が大きいフレームは波形の周期性が高く波形の類似性が

高いため、そのフレームを伸縮処理しても劣化は検知され難いと仮定し、速度変換処理のために伸縮すべきフレームとして、定常性の大きいフレームから優先的に所定数Ns個の選択を行う。そして、データ伸縮制御手段３００６は、伸縮すべきフレームとして選択されたフレームかどうかを判断し、データ伸縮手段３００２が伸縮すべきかどうかを制御する。その結果、入力された符号化データは、フレーム復号化手段３００１で１フレーム単位で復号化され、データ伸縮制御手段によって伸縮すべきと判断されたフレームについて、波形の伸縮を行い、それ以外のフレームについては、そのまま出力を行う。このように、あらかじめ定常性演算手段で求めた音声の定常性を用いて、フレーム選択手段でフレームサイクル内で伸縮すべき最適なフレームを求めておくことにより、速度変換処理音声として、波形の伸縮による処理劣化が検知され難くいものとすることが可能となる。

【００９９】なお、本実施の形態５では、各フレームにおける定常性を示す値として、正規化自己相関関数を利用しているが、例えば、周波数スペクトルの変化度合などを用いることも有効であると考えられる。

【０１００】（実施の形態６）以下、本発明の第６の実施の形態について、図面を参照しながら説明する。図１２は本発明の第６の実施の形態における音声再生装置のブロック図を示すものである。図１２において、３００１はフレーム復号化手段、３００２はデータ伸縮手段、３００３は伸縮頻度制御手段、６００４はエネルギー変化度合演算手段、６００５はフレーム選択手段、３００６はデータ伸縮制御手段である。以下に、その動作について説明する。

【０１０１】本実施の形態６は、フレーム単位で復号化処理を行う音声に対して速度変換処理を施す音声再生装置の一例を示すものである。図１２において、フレーム復号化手段３００１、データ伸縮手段３００２、伸縮頻度制御手段３００３、データ伸縮制御手段３００６は、実施の形態３と同様の動作を行うものである。本実施の形態３では、伸縮処理を行うべきフレームの選択を行うフレーム選択手段の働きを中心に説明を行う。

【０１０２】本実施の形態６では、音声波形のエネルギー変化度合に着目する。ここでは、１フレーム内をさらに複数の小区間に分割した各小区間でのエネルギー値を求め、各小区間の前値との差分値を求めることにより、エネルギーの変化度合を求める。そして、このエネルギーの時間的な変化度合を継続的に監視することによって、時間的に継続する区間に対するマスキング効果である，継時マスキング（temporal masking）の影響を考慮した処理フレームの選択を行う。このマスキングに関しては、参考文献１：Moore の本、に詳しく記述されているが、マスカーの前後の双方の区間に対してマスキング効果を生じ、この性質を利用すれば、時間伸縮処理による劣化を検知され難くできる。すなわち、大きなエネルギーのフレームの直後の小さなエネルギーのフレームは、マスク(backward masking)され、時間軸伸縮の劣化が検知され難い。あるいは、小さなエネルギーのフレームに継続して直後に大きなエネルギーのフレームが到来する場合、前の小さいエネルギーのフレームは、マスク（forward masking）され、時間伸縮処理の劣化は、検知されにくい。また、これらのマスキング量は、マスカーとのレベル差、および時間差によって値が異なっている。ただし、高速再生時における時間軸圧縮処理により新たに発生する継時マスキング効果によって、新たにエネルギーの小さい部分の聴き取りが困難になる，ということがないように注意する必要がある。

【０１０３】エネルギー変化度合演算手段６００４では、伸縮頻度制御手段で決定されたフレームサイクル数分のエネルギー変化度合を予め求める。次に、フレーム選択手段６００５は、先に求められたNf個のエネルギー変化度合の値を参考に、継時マスキング効果による処理劣化が検知されにくいフレームから優先的に所定数Ns個の選択を行う。その際、時間軸圧縮を行うことにより、エネルギーの小さい区間の聴き取りが困難になる，ということがないように注意しなければならない。すなわち、エネルギーの大きいフレームに挟まれたエネルギーの小さいフレームは、時間長が短くなることによる、前方・後方マスキング効果の増大が予想されるため、ほかのフレームを選択するようにする。そして、データ伸縮制御手段３００６は、当該フレームが伸縮すべきフレームとして選択されたフレームかどうかを判断し、データ伸縮手段３００２が伸縮すべきかどうかを制御する。その結果、入力された符号化データは、フレーム復号化手段３００１で１フレーム単位で復号化され、データ伸縮制御手段によって伸縮すべきと判断されたフレームについて、波形の伸縮を行い、それ以外のフレームについては、そのまま出力を行う。このように、あらかじめエネルギー変化度合演算手段で求めたエネルギーの変化度合を用いて、フレーム選択手段でフレームサイクル内で伸縮すべき最適なフレームを求めておくことにより、速度変換処理音声として、波形の伸縮による処理劣化が検知され難くいものとすることが可能となる。

【０１０４】なお、本実施の形態６では、各フレームにおけるエネルギー変化度合を示す値を指標として継時マスキング効果を利用しているが、例えば１フレーム内をさらに複数の小区間に分割した各小区間ごとの平均振幅値を求め、各小区間の前値との差分値を求めることにより、平均振幅の変化度合を代用して用いることも有効であると考えられる。

【０１０５】（実施の形態７）以下、本発明の第７の実施の形態について、図面を参照しながら説明する。図１３は本発明の第７の実施の形態による音声再生装置のブロック図を示すものである。図１３において、３００１はフレーム復号化手段、３００２はデータ伸縮手段、３

００３は伸縮頻度制御手段、４００４は音声らしさ演算手段、５００４は定常性演算手段、６００４はエネルギー変化度合演算手段、７００５はフレーム選択手段、３００６はデータ伸縮制御手段である。以下に、その動作について説明する。

【０１０６】本実施の形態７は、フレーム単位で復号化処理を行う音声に対して速度変換処理を施す音声再生装置の一例を示すものである。図１３において、フレーム復号化手段３００１、データ伸縮手段３００２、伸縮頻度制御手段３００３、データ伸縮制御手段３００６は、実施の形態３と同様の動作を行うものである。また、音声らしさ演算手段４００４は、実施の形態４と、定常性演算手段５００４は、実施の形態5 と、エネルギー変化度合演算手段６００４は、実施の形態6 と同様の動作を行う。本実施の形態７では、伸縮処理を行うべきフレームの選択を行うフレーム選択手段７００５の働きを中心に説明を行う。

【０１０７】速度変換処理によって処理された音声から得るべき情報は、音声言語情報であると仮定すると、対象とする音声が処理により加工されたことによって、聴取者の了解性が低下することは望ましくない。あるいは、速度変換処理を適用することによって了解性を高めることができる可能性があることは、学会発表等より明らかにされつつある（参考文献3 ）,4）） 。例えば、音声聴取の際の時間処理能力が低下している高齢者においては、速度を低下させることによって、了解性が高められることが確認されている。本実施の形態７では、速度変換処理によって了解性を向上させ、処理による劣化を最小限に抑える、あるいは、速度変換処理によって自然性が劣化せず効率的に音声情報を聴取しやすくする、の２つの処理形態を提供するものである。フレーム選択手段７００５は、音声らしさ演算手段の出力結果と、定常性演算手段の出力結果と、エネルギー変換度合演算手段によって得られるマスキング条件とをもとに、各フレームに対する分析結果を数値化し、これをもとに、自然性を重視した場合、了解性を重視した場合、の双方に関して、選択すべきフレームを決定するものである。

【０１０８】まず、自然性の劣化を少なく、効率的に聴取する場合の処理を説明する。この場合は、音声らしさ演算手段によって得られた非音声区間のフレームに対する優先度を大きくする。そして、残りの２つの分析結果を考慮して、最終的なフレーム選択を決定する。

【０１０９】次に、了解性を高め、聴き取りやすい音声を得る場合の処理を説明する。この場合は、エネルギーの小さい子音語頭部が継時マスキングされないようにエネルギー変化度合のパラメータの優先度を高くする。そして、残りの２つの分析結果を考慮して、最終的なフレーム選択を決定する。

【０１１０】このように、あらかじめエネルギー変化度合演算手段で求めたエネルギーの変化度合を用いて、（あるいは、音声らしさ演算手段によって得られた非音声区間のフレームに対する優先度を大きくして、）フレーム選択手段でフレームサイクル内で伸縮すべき最適なフレームを求めておくことにより、速度変換処理音声として、自然性・了解性の優先度合いを考慮した波形の伸縮を行うことができるものである。

【０１１１】なお、本実施の形態７では、請求項9 に対応するうちの一例として、エネルギー演算手段、音声らしさ演算手段、定常性演算手段、エネルギー変化度合演算手段の4 つの手段のうち、後者の３つを備えたものを説明したが、エネルギー演算手段を判定条件に加えてどのフレームに対して時間軸伸縮を加えるべきかを、より厳密に推定することも可能である。本発明では、これは4 つの演算手段のうち２つ以上を備えて総合的な推定を行うことで、再生音の聴取条件などに関して複数の選択肢を与えるものである。

【０１１２】（実施の形態８）以下、本発明の実施の形態８について、図面を参照しながら説明する。まず、以下の実施の形態８～１１の説明に先立ち、MPEG1 オーディオレイヤ１/ ２符号化方式について説明する。MPEG1 オーディオレイヤ１/ ２符号化方式は、図２６に示すブロック図で表される。１６ビット直線量子化された入力信号は、サブバンド分析フィルタで３２帯域のサブバンド信号に分割される。フィルタは、５１２タップPFB（Polyphase Filter Bank ）で実現される。各サブバンド信号に対してスケールファクタを計算し、ダイナミックレンジを揃える。スケールファクタの計算は、レイヤ１では各帯域１２サンプルごと、すなわち全体で３８４サンプルごとに、レイヤ２ではその３倍の１１５２サンプルを１ブロックとして３８４サンプルごとに行われる。このため、レイヤ２では解像度が増し、符号化品質が向上する。しかし、このままではレイヤ２のスケールファクタの数はレイヤ１の３倍になり、圧縮率の低下を招く。そこで、レイヤ２では３つのスケールファクタの組み合わせに応じて１つの新たな値（スケールファクタ選択情報）を割り当てて表現し、圧縮率低下を防ぐ。

【０１１３】図１４は本発明の実施の形態８における音声再生装置のブロック図を示すものである。図１４において、１０１はフレーム逆パッキング手段、１０２は逆量子化手段、１０３はデータ伸縮手段、１０４はサブバンド合成フィルタ手段、１０６はフレームカウント手段、12-1-1はエネルギー演算手段、12-1-2は伸縮頻度制御手段、12-1-3はフレーム選択手段、１０７はデータ伸縮制御手段である。

【０１１４】図１５は、本発明の実施の形態８における、エネルギー演算手段12-1-1がフレームのエネルギーを推定する過程を示すフローチャートである。以下に、その動作について説明する。

【０１１５】本実施の形態８は、ＭＰＥＧ１オーディオレイヤ２のビットストリームをデコードする際の中間デ

ータに対して速度変換処理を施す音声再生装置の例を示すものである。ＭＰＥＧ１オーディオレイヤ２のビットストリームは、ヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などから成り立っている。

【０１１６】図１４において、入力されたＭＰＥＧ１オーディオレイヤ２のビットストリームは、フレーム逆パッキング手段１０１によって、当該ビットストリームからヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などの個々の情報に分離される。

【０１１７】ここで、スケールファクタインデックスは、再生時の波形倍率を示し、各チャンネル、各有効サブバンド、各ブロックごとに存在する。スケールファクタインデックスは０から６２までの値を取り、０が最もエネルギーが大きく、６２が最もエネルギーが小さい。ただしビット割当情報が０の場合はスケールファクタインデックスは存在しない。また、ビット割当情報は、エンコード時に割当てるべきビット数に関連した値で、各チャンネル、各有効サブバンドごとに存在する。

【０１１８】既に、述べたことでもあるが、ＭＰＥＧ１オーディオレイヤ２におけるチャンネルは、右チャンネルと左チャンネルの２チャンネル存在しうる。また、ＭＰＥＧ１オーディオレイヤ２におけるサブバンドは、全帯域を３２等分割したものであり、周波数の低い順に、第０サブバンド、第１サブバンド、第２サブバンドから第３１サブバンドまで存在する。

【０１１９】ここで、サブバンドについては、サンプリング周波数が３２ｋＨｚの場合、０～１６０００Ｈｚの帯域を３２等分割するため、一つのサブバンドは５００Ｈｚの幅を持つ。ただし、３２個のサブバンドのうち有効なサブバンド数が制限される。例えば１９２ｋｂｐｓステレオの場合、０～３１の３２個のサブバンドのうち、０～２９までの３０個のサブバンドを有効サブバンドとするため、第３０、第３１サブバンドのビット割当情報や、スケールファクタインデックスは存在しない。この時、周波数帯域は０～１５０００Ｈｚとなる（１６０００÷３２×３０＝１５０００より）。

【０１２０】また、ＭＰＥＧ１オーディオレイヤ２におけるブロックとは、フレームを時間領域で３等分割した領域であり、時間順に第０ブロック、第１ブロック、第２ブロックまで存在する。サンプリング周波数が３２ｋＨｚの場合、１ブロック長＝１２ｍｓである。１フレーム長は、サンプリング周波数が３２ｋＨｚの場合３６ｍｓである。

【０１２１】エネルギー演算手段12-1-1は、第０ブロックの第０サブバンドの左チャンネルのスケールファクタインデックスｓｃｆ＿Ｌ０と、第０ブロックの第０サブバンドの右チャンネルのスケールファクタインデックスｓｃｆ＿Ｒ０とを用いて、フレームサイクル内の各フレームナンバｆｒｍに対するエネルギー推定値ｅ［ｆｒｍ］を求める。より詳しくは、スケールファクタインデックスの小さいフレームほどエネルギーは大きいもので、上記ｓｃｆ＿Ｌ０とｓｃｆ＿Ｒ０のうちどちらか小さい方の値を用いて、上記エネルギー推定値ｅ［ｆｒｍ］を求める。

【０１２２】ｓｃｆ＿Ｌ０とｓｃｆ＿Ｒ０の一方が存在しない時は、エネルギー演算手段12-1-1は、存在するもう一方の値を用いて、エネルギー推定値ｅ［ｆｒｍ］を求める。ｓｃｆ＿Ｌ０とｓｃｆ＿Ｒ０の両方が存在しない時は、エネルギー演算手段12-1-1は、速度変換フレーム選択候補の優先順位を最低にすることを意味する所定値を、エネルギー推定値ｅ［ｆｒｍ］に代入する。

【０１２３】伸縮頻度制御手段12-1-2は、与えられた速度比に応じて、フレームサイクル数と、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する。例えば０．９倍速の時、９フレームのうち２フレームを速度変換を施すフレームとして選択する。つまりフレームサイクル数は９であり、フレームナンバｆｒｍは０から８を変動する。フレーム選択手段12-1-3は、エネルギー演算手段12-1-1が出力するフレームサイクル中の全フレームに対するエネルギー推定値ｅ［ｆｒｍ］の小さいフレームから順に、伸縮処理を行うフレームを選択する。ｅ［ｆｒｍ］の小さいフレームを優先的に選択すれば、エネルギーの小さい音の部分が速度変換処理されることになる。

【０１２４】なお、第０ブロックの第０サブバンドの左チャンネルのスケールファクタインデックスｓｃｆ＿Ｌ０と、第１ブロックの第０サブバンドの左チャンネルのスケールファクタインデックスｓｃｆ＿Ｌ１と、第２ブロックの第０サブバンドの左チャンネルのスケールファクタインデックスｓｃｆ＿Ｌ２と、第０ブロックの第０サブバンドの右チャンネルのスケールファクタインデックスｓｃｆ＿Ｒ０と、第１ブロックの第０サブバンドの右チャンネルのスケールファクタインデックスｓｃｆ＿Ｒ１と、第２ブロックの第０サブバンドの右チャンネルのスケールファクタインデックスｓｃｆ＿Ｒ２とのうちの最小値を用いて、エネルギー推定値ｅ［ｆｒｍ］を求めるようにしてもよい。

【０１２５】以上のように、本実施の形態８によれば、エネルギー演算手段12-1-1は、再生時の波形倍率を示すスケールファクタインデックスの値をもとに、音声信号のエネルギーを推定するようにし、その結果に応じて速度変換を施すフレームを選択するようにしたので、ＭＰＥＧデコード後のＰＣＭデータのエネルギー演算が不要となり、ＭＰＥＧ１オーディオレイヤ２のビットストリームをデコードする際の中間データに対して、速度変換フレーム選択、及び速度変換処理を施すことが可能となるため、少ない演算量で速度変換処理を実現することができるものである。

【０１２６】（実施の形態９）以下、本発明の実施の形態９について、図面を参照しながら説明する。図１６は、本発明の実施の形態９における音声再生装置のブロック図を示すものである。図１６において、１０１はフレーム逆パッキング手段、１０２は逆量子化手段、１０３はデータ伸縮手段、１０４はサブバンド合成フィルタ手段、１０６はフレームカウント手段、13-1-1は定常性演算手段、12-1-2は伸縮頻度制御手段、13-1-3はフレーム選択手段、１０７はデータ伸縮制御手段である。表４は、本発明の実施の形態９において定常性演算手段13-1-1が出力する、定常性検出による速度変換フレーム選択優先順位である。以下に、その動作について説明する。

【０１２７】

【表４】

ord [frm]	scfsi_L0	scfsi_R0
1	2	2
2	1	2
2	2	1
2	2	3
2	3	2
3	1	1
3	1	3
3	3	1
3	3	3
4	0	2
4	2	0
5	0	1
5	1	0
5	0	3
5	3	0
6	0	0

【０１２８】本実施の形態９は、ＭＰＥＧ１オーディオレイヤ２のビットストリームをデコードする際の中間データに対して速度変換処理を施す音声再生装置の例を示すものである。ＭＰＥＧ１オーディオレイヤ２のビットストリームは、ヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などから成り立っている。

【０１２９】図１６において、入力されたＭＰＥＧ１オーディオレイヤ２のビットストリームは、フレーム逆パッキング手段１０１によって、当該ビットストリームからヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などの個々の情報に分離される。スケールファクタ選択情報は、波形定常性を示すものであり、各チャンネル、各有効サブバンドごとに存在している。スケールファクタ選択情報は、０、１、２、３の値を取りうる。スケールファクタ選択情報が０のとき最も定常性が低く、２のとき最も定常性が高いものと見なす。スケールファクタ選択情報が１および３のとき定常性は同等であると見なす。

【０１３０】定常性演算手段13-1-1は、第０サブバンドの左チャンネルのスケールファクタ選択情報ｓｃｆｓｉ＿Ｌ０と、第０サブバンドの右チャンネルのスケールファクタ選択情報ｓｃｆｓｉ＿Ｒ０とを用いて、フレームサイクル内の各フレームナンバｆｒｍに対する速度変換フレーム選択優先順位ｏｒｄ［ｆｒｍ］を求める。定常性演算手段13-1-1は、フレームサイクル内の全フレームのｏｒｄ［ｆｒｍ］を、表４に示す規則に従って求める。ｓｃｆｓｉ＿Ｌ０とｓｃｆｓｉ＿Ｒ０のどちらか一つまたは両方が存在しないときは、定常性演算手段13-1-1は、速度変換フレーム選択候補の優先順位を最低にすることを意味する所定値を、速度変換フレーム選択優先順位ｏｒｄ［ｆｒｍ］に代入する。

【０１３１】伸縮頻度制御手段12-1-2は、与えられた速度比に応じて、フレームサイクル数とそのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する。フレーム選択手段13-1-3は、定常性演算手段13-1-1が出力するフレームサイクル中の全フレームに対する速度変換フレーム選択優先順位ｏｒｄ［ｆｒｍ］の高いフレームから順に、伸縮処理を行うフレームを選択する。

【０１３２】以上のように、本実施の形態９によれば、定常性演算手段13-1-1は、波形定常性を示すスケールファクタ選択情報の値をもとに、音声信号の定常性を推定することにより、ＭＰＥＧデコード後のＰＣＭデータの定常性演算が不要となり、ＭＰＥＧ１オーディオレイヤ２のビットストリームをデコードする際の中間データに対して速度変換フレーム選択、及び速度変換処理を施すことが可能となるため、少ない演算量で速度変換処理実現することができるものである。

【０１３３】このような、本実施の形態９では、速度変換による音質劣化の少ない定常性の高いフレームを選択して速度変換するというところに特徴があり、このように、話速変換ができるので、語学学習に適しているものであり、また、定常性演算処理が不要となるため、演算量を削減できる、という特徴をも有するものである。

【０１３４】（実施の形態１０）以下、本発明の実施の形態１０について、図面を参照しながら説明する。図１７は、本発明の実施の形態１０における音声再生装置のブロック図を示すものである。図１７において、１０１はフレーム逆パッキング手段、１０２は逆量子化手段、１０３はデータ伸縮手段、１０４はサブバンド合成フィルタ手段、１０６はフレームカウント手段、14-1-1はエネルギー変化度合演算手段、12-1-2は伸縮頻度制御手段、14-1-3はフレーム選択手段、１０７はデータ伸縮制御手段である。以下に、その動作について説明する。

【０１３５】本実施の形態１０は、ＭＰＥＧ１オーディオレイヤ２のビットストリームをデコードする際の中間データに対して速度変換処理を施す音声再生装置の例を示すものである。ＭＰＥＧ１オーディオレイヤ２のビットストリームは、ヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などから成り立っている。

【０１３６】図１７において、入力されたＭＰＥＧ１オーディオレイヤ２のビットストリームは、フレーム逆パ

ッキング手段１０１によって、当該ビットストリームからヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などの個々の情報に分離される。

【０１３７】エネルギー変化度合演算手段14-1-1は、第０ブロックの第０サブバンドの左チャンネルのスケールファクタインデックスｓｃｆ＿Ｌ０と、第１ブロックの第０サブバンドの左チャンネルのスケールファクタインデックスｓｃｆ＿Ｌ１と、第２ブロックの第０サブバンドの左チャンネルのスケールファクタインデックスｓｃｆ＿Ｌ２と、第０ブロックの第０サブバンドの右チャンネルのスケールファクタインデックスｓｃｆ＿Ｒ０と、第１ブロックの第０サブバンドの右チャンネルのスケールファクタインデックスｓｃｆ＿Ｒ１と、第２ブロックの第０サブバンドの右チャンネルのスケールファクタインデックスｓｃｆ＿Ｒ２とを用いて、フレームサイクル内と、フレームサイクルの前後１フレーム、の各フレームナンバｆｒｍに対する各チャンネルの各ブロックのエネルギー推定値ｅ［ｃｈ］［ｂｌｋ］［ｆｒｍ］を求める。フレームサイクル内とフレームサイクルの前後１フレームとは、例えばフレームサイクル数が９の場合、９フレームの前後１フレームということで、１１フレームとなる。

【０１３８】即ち、各ブロックの第０サブバンドの各チャンネルのスケールファクタインデックスに対応する、各フレームナンバの各チャンネルの各ブロックのエネルギー推定値ｅ［ｃｈ］［ｂｌｋ］［ｆｒｍ］を、フレームサイクル内とフレームサイクルの前後１フレームについて求める。スケールファクタインデックスの小さいブロックほどエネルギーは大きい。

【０１３９】また、スケールファクタインデックスが存在しないとき、エネルギーは０である。即ち、ｓｃｆ＿Ｌ０が存在しないフレームの場合、ｅ［０］［０］［ｆｒｍ］＝０とする。ｓｃｆ＿Ｌ１が存在しないフレームの場合、ｅ［０］［１］［ｆｒｍ］＝０とする。ｓｃｆ＿Ｌ２が存在しないフレームの場合、ｅ［０］［２］［ｆｒｍ］＝０とする。ｓｃｆ＿Ｒ０が存在しないフレームの場合、ｅ［１］［０］［ｆｒｍ］＝０とする。ｓｃｆ＿Ｒ１が存在しないフレームの場合、ｅ［１］［１］［ｆｒｍ］＝０とする。ｓｃｆ＿Ｒ２が存在しないフレームの場合、ｅ［１］［２］［ｆｒｍ］＝０とする。

【０１４０】次に、フレームサイクル内の各フレームナンバｆｒｍに対するエネルギー推定値ｅ［ｃｈ］［ｂｌｋ］［ｆｒｍ］のブロック内の最大値ｅｍａｘ［ｃｈ］［ｆｒｍ］を、全フレームサイクルについて求める。フレームサイクルの前後１フレームのｅｍａｘ［ｃｈ］［ｆｒｍ］は、求めなくてよい。

【０１４１】次に、フレームサイクル内の各フレームナンバｆｒｍに対して、エネルギー推定値ｅ［０］［２］［ｆｒｍ－１］－ｅｍａｘ［０］［ｆｒｍ］と、エネルギー推定値ｅ［１］［２］［ｆｒｍ－１］－ｅｍａｘ［１］［ｆｒｍ］と、エネルギー推定値ｅ［０］［０］［ｆｒｍ＋１］－ｅｍａｘ［０］［ｆｒｍ］と、エネルギー推定値ｅ［１］［０］［ｆｒｍ＋１］－ｅｍａｘ［１］［ｆｒｍ］の４個の値を求め、４個の値のうちの最大値を、速度変換フレーム選択優先度ｐ［ｆｒｍ］に代入する。

【０１４２】伸縮頻度制御手段12-1-2は、与えられた速度比に応じて、フレームサイクル数と、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する。フレーム選択手段14-1-3は、エネルギー変化度合演算手段14-1-1が出力するフレームサイクル中の全フレームに対する速度変換フレーム選択優先度ｐ［ｆｒｍ］の大きいフレームから順に、伸縮処理を行うフレームを選択する。速度変換フレーム選択優先度ｐ［ｆｒｍ］の大きいフレームほど、非同時マスキングでマスキングされやすいので、速度変換による音質劣化が知覚されにくいことが特徴となる。非同時マスキングについては、Ｂ．Ｃ．Ｊ．ムーア著、誠信書房発行、聴覚心理学概論に詳しく記述されている。

【０１４３】以上のように、本実施の形態１０によれば、エネルギー変化度合演算手段14-1-1は、再生時の波形倍率を示すスケールファクタインデックスの値をもとに、音声信号のエネルギー変化度合を推定し、速度変換フレーム選択優先度ｐ［ｆｒｍ］の大きいフレームを優先的に速度変換するようにしたので、ＭＰＥＧデコード後のＰＣＭデータのエネルギー変化度合演算が不要となり、ＭＰＥＧ１オーディオレイヤ２のビットストリームをデコードする際の中間データに対して速度変換フレーム選択、及び速度変換処理を施すことが可能となるため、少ない演算量で実現できることが特徴である。また、この方法は、話速変換ができるので、語学学習に適した音声処理を行うことができる。

【０１４４】（実施の形態１１）以下、本発明の実施の形態１１について、図面を参照しながら説明する。図１８は、本発明の実施の形態１１における音声再生装置のブロック図を示すものである。図１８において、１０１はフレーム逆パッキング手段、１０２は逆量子化手段、１０３はデータ伸縮手段、１０４はサブバンド合成フィルタ手段、１０６はフレームカウント手段、12-1-1はエネルギー演算手段、13-1-1は定常性演算手段、14-1-1はエネルギー変化度合演算手段、12-1-2は伸縮頻度制御手段、15-1-3はフレーム選択手段、１０７はデータ伸縮制御手段である。以下に、その動作について説明する。

【０１４５】本実施の形態１１は、ＭＰＥＧ１オーディオレイヤ２のビットストリームをデコードする際の中間データに対して速度変換処理を施す音声再生装置の例を示すものである。ＭＰＥＧ１オーディオレイヤ２のビットストリームは、ヘッダ、ビット割当情報、スケールフ

ァクタインデックス、スケールファクタ選択情報、サンプルデータ情報などから成り立っている。

【０１４６】図１８において、入力されたＭＰＥＧ１オーディオレイヤ２のビットストリームは、フレーム逆パッキング手段１０１によって、当該ビットストリームからヘッダ、ビット割当情報、スケールファクタインデックス、スケールファクタ選択情報、サンプルデータ情報などの個々の情報に分離される。

【０１４７】エネルギー演算手段12-1-1は、本発明の実施の形態８に記述した方法で、フレームサイクル内の各フレームナンバｆｒｍに対するエネルギー推定値ｅ［ｆｒｍ］を求める。

【０１４８】定常性演算手段13-1-1は、本発明の実施の形態９に記述した方法で、フレームサイクル内の各フレームナンバｆｒｍに対する速度変換フレーム選択優先順位ｏｒｄ［ｆｒｍ］を求める。

【０１４９】エネルギー変化度合演算手段14-1-1は、本発明の実施の形態１０に記述した方法で、フレームサイクル内の各フレームナンバｆｒｍに対する速度変換フレーム選択優先度ｐ［ｆｒｍ］を求める。

【０１５０】伸縮頻度制御手段12-1-2は、与えられた速度比に応じて、フレームサイクル数と、そのフレームサイクル数内で伸縮処理を行うフレーム数とを設定する。自然性の劣化を少なく、効率的に聴取したい場合、フレーム選択手段15-1-3は、エネルギー演算手段12-1-1が出力するフレームサイクル中の全フレームに対するエネルギー推定値ｅ［ｆｒｍ］の小さいフレームから順に、伸縮処理を行うフレームを選択する。了解性を高め、聴き取りやすい音声を得たい場合、フレーム選択手段15-1-3は、定常性演算手段13-1-1が出力するフレームサイクル中の全フレームに対する速度変換フレーム選択優先順位ｏｒｄ［ｆｒｍ］の高いフレームから順に、伸縮処理を行うフレームを選択する。このときフレームサイクル内の速度変換フレーム選択優先順位ｏｒｄ［ｆｒｍ］の値が同一で優先順位がつけられない場合は、エネルギー変化度合演算手段14-1-1が出力する速度変換フレーム選択優先度ｐ［ｆｒｍ］を用いて、そのｐ［ｆｒｍ］の大きいフレームを優先的に選択するようにして速度変換フレーム選択優先順位ｏｒｄ［ｆｒｍ］の値が同一なフレームに対して細分化した優先順位をつける。

【０１５１】以上のように、本実施の形態１１によれば、エネルギー演算手段12-1-1と、定常性演算手段13-1-1と、エネルギー変化度合演算手段14-1-1は、再生時の波形倍率を示すスケールファクタインデックスと、スケールファクタ選択情報の値をもとに、音声信号のエネルギーと、定常性と、エネルギー変化度合を推定し、自然性重視の場合、ｅ［ｆｒｍ］の小さいフレームを選択し、了解性重視の場合、ｏｒｄ［ｆｒｍ］の小さいフレームを選択し、ｏｒｄ［ｆｒｍ］が同一の値の場合、ｐ［ｆｒｍ］の大きいフレームを優先的に選択するようにしたので、ＭＰＥＧデコード後のＰＣＭデータのエネルギーと、定常性と、エネルギー変化度合の演算が不要となり、ＭＰＥＧ１オーディオレイヤ２のビットストリームをデコードする際の中間データに対して速度変換フレーム選択、及び速度変換処理を施すことが可能となるため、少ない演算量で所望の話速変換処理を行うことができる効果が得られる。

【０１５２】なお、本発明の実施の形態４に記載されている音声らしさ演算手段４００４が本実施の形態１１に記載されていないのは、ＭＰＥＧ１オーディオレイヤ２のビットストリームに音声らしさを示す情報が含まれていないためである。

【０１５３】［参考文献］

1) 鈴木，三崎，“高品質速度変換方式のDSP による実現”，信学技報，SP90-34(1990)

2) 比企他，“連続音声中の音韻区分の持続時間の性質”，信学誌，第50巻5号，pp. 849-856(1967)

3) 中山，三 ，“日本人学習者に対する英語の語頭強調処理による受聴明瞭度の改善”，音講論集，1-8-21（1998. 3)

4) 細井，目方他，“補聴効果評価のための67-S早口語音聴力検査”，Audiology Japan, vol. 36. No. 5, pp. 299-300(1993)

5) B. C. J. Moore 著（大串健吾監訳），“聴覚心理学概論”誠信書房（非同期マスキングに関しても参照）

【０１５４】

【発明の効果】請求項１にかかる音声再生装置によれば、音声復号化手段、選択手段、フレームシーケンステーブル、フレームカウント手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、音声復号化手段は、入力される音声信号をフレーム単位で復号し、選択手段は、与えられる速度比に対応したフレームシーケンスをフレームシーケンステーブルへ出力すると共に、該フレームシーケンスのフレームサイクル数をフレームカウント手段へ出力し、フレームシーケンステーブルは、選択手段からのフレームシーケンスを記憶し、フレームカウント手段は、フレームサイクル数に基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、データ伸縮制御手段は、フレームカウント手段のカウント値に対応したフレームシーケンステーブルのフレームシーケンスを参照して、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行うことを特徴とするものとしたので、フレーム内データで完結する一定速度比の時間軸圧縮処理または時間軸伸長処理を基本とした簡素な構成によって、所望の速度比（再生速度）にて高品質な速度変換処理を実現する音声再生装置を提供

することができる効果がある。

【０１５５】また、請求項２にかかる音声再生装置によれば、請求項１に記載の音声再生装置において、音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された音声信号を復号することを特徴とするものとしたので、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化されたデータに対して、処理劣化の少ない速度変換処理を行うことができる音声再生装置を提供できる効果がある。

【０１５６】また、請求項３にかかる音声再生装置によれば、請求項１に記載の音声再生装置において、フレームシーケンスは、連続する時間軸圧縮フレームのフレーム数と、連続する時間軸処理無しフレームのフレーム数がいずれも最小となるよう配置されたことを特徴とするものとしたので、フレーム内データで完結する一定速度比の時間軸圧縮または時間軸伸長処理を基本とした簡素な構成によって、所望の速度比（再生速度）にて高品質な速度変換処理を実現する音声再生装置を提供することができる効果がある。

【０１５７】また、請求項４にかかる音声再生装置によれば、請求項１に記載の音声再生装置において、フレームシーケンスは、連続する時間軸伸長フレームのフレーム数と、連続する時間軸処理無しフレームのフレーム数がいずれも最小となるよう配置されたことを特徴とするものとしたので、フレーム内データで完結する一定速度比の時間軸圧縮または時間軸伸長処理を基本とした簡素な構成によって、所望の速度比（再生速度）にて高品質な速度変換処理を実現する音声再生装置を提供することができる効果がある。

【０１５８】また、請求項５にかかる音声再生装置によれば、音声復号化手段、伸縮頻度制御手段、フレームカウント手段、エネルギー演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された符号化音声信号を復号し、伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、エネルギー演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギーを推定し、フレーム選択手段は、フレームサイクル数Ｎｆのフレーム内でエネルギーの小さいフレームからＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行うことを特徴とするものとしたので、エネルギーの小さいフレームでの時間軸伸縮は処理劣化が検知され難いことを利用し、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化されたデータに対し、エネルギーの小さいフレームを優先的に選択することができ、高品質な速度変換処理音声を得ることができる音声再生装置を提供することができる効果がある。

【０１５９】また、請求項６にかかる音声再生装置によれば、音声復号化手段、伸縮頻度制御手段、フレームカウント手段、定常性演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された音声信号を復号し、伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、定常性演算手段は、符号化音声信号のスケールファクタ選択情報をもとにフレームサイクル数Ｎｆ分の符号化音声信号の定常性を推定し、フレーム選択手段は、フレームサイクル数Ｎｆのフレーム内での定常性の高いフレームからＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行うことを特徴とする音声再生装置としたので、定常性の高いフレームでは重み付け加算法による劣化が検知され難いことを利用し、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化されたデータに対し、定常性の高いフレームを優先的に選択することができ、高品質な速度変換処理音声を得ることができる音声再生装置を提供することができる効果がある。

【０１６０】また、請求項７にかかる音声再生装置によれば、音声復号化手段、伸縮頻度制御手段、フレームカウント手段、エネルギー変化度合演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された音声信号を復号し、伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号

化手段で処理する符号化音声信号のフレーム数をカウントし、エネルギー変化度合演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギー変化度合を推定し、フレーム選択手段は、フレームサイクル数Ｎｆのフレーム内でエネルギー変化度合に基づき継時マスキング効果による処理劣化が少ないフレームからＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行うことを特徴とするものとしたので、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化されたデータに対し、エネルギー変化度合に基づき処理劣化が継時マスキング効果によって検知しにくいフレームを選択することとなり、定常性の高いフレームを優先的に選択することができ、高品質な速度変換処理音声を得ることができる音声再生装置を提供することができる効果がある。

【０１６１】また、請求項８にかかる音声再生装置によれば、音声復号化手段、伸縮頻度制御手段、フレームカウント手段、演算手段、フレーム選択手段、データ伸縮制御手段、データ伸縮手段を備える音声再生装置であって、音声復号化手段は、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された符号化音声信号を復号し、伸縮頻度制御手段は、与えられる速度比に応じた、フレームサイクル数Ｎｆ、時間軸圧縮または時間軸伸長するフレーム数Ｎｓを設定し、フレームカウント手段は、フレームサイクル数Ｎｆに基づいて音声復号化手段で処理する符号化音声信号のフレーム数をカウントし、演算手段は、エネルギー演算手段、定常性演算手段、エネルギー変化度合演算手段のいずれか２つ以上を備え、エネルギー演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギーを推定し、定常性演算手段は、符号化音声信号のスケールファクタ選択情報をもとにフレームサイクル数Ｎｆ分の符号化音声信号の定常性を推定し、エネルギー変化度合演算手段は、符号化音声信号のスケールファクタインデックスをもとにフレームサイクル数Ｎｆ分の符号化音声信号のエネルギー変化度合を推定し、フレーム選択手段は、演算手段の出力をもとにＮｓ個のフレームを時間軸圧縮または時間軸伸長するフレームとして決定し、データ伸縮制御手段は、フレームカウント手段のカウント値とフレーム選択手段の決定に基づき、音声復号化手段から出力されるフレームを時間軸圧縮もしくは時間軸伸長、または時間軸変換なしのどちらで処理するかをデータ伸縮手段に指定し、データ伸縮手段は、データ伸縮制御手段の指定に基づいて音声復号化手段から出力されるフレームに対して時間軸変換処理を行うことを特徴とするものとしたので、ＭＰＥＧ１オーディオレイヤ２符号化方式にて符号化された符号化音声信号に対し、上記複数の演算手段の出力を総合的に判断して選択すべきフレームを決定でき、目的に応じてそれぞれ高品質な速度変換処理音声を得ることができる音声再生装置を提供することができる効果がある。

【０１６２】また、請求項９にかかる音声再生装置によれば、請求項１～８のいずれかに記載の音声再生装置において、データ伸縮手段は、クロスフェード手段を備え、クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを重み付け加算することを特徴とするものとしたので、フレーム内データで完結する一定速度比の時間軸圧縮または時間軸伸長処理を基本とした簡素な構成によって、所望の速度比（再生速度）にて高品質な速度変換処理を行なうことができる音声再生装置を提供することができる効果がある。

【０１６３】また、請求項１０にかかる音声再生装置によれば、請求項１～８のいずれかに記載の音声再生装置において、データ伸縮手段は、相関演算手段、クロスフェード手段を備え、相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、セグメント間の相関値を演算し、相関値が高くなる位置で重み付け加算するためのシフト量を決定し、クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算することを特徴とするものとしたので、フレームを構成するセグメント間の相関が高くなる位置に波形データをシフトさせて相関演算を行い、かつ各時間軸圧縮または時間軸伸長の処理において上記シフト量を考慮した処理を行うことによって、重み付け加算するフレームの位相の整合性を高められるため、音声信号の処理劣化の少ない速度変換処理を行うことができる音声再生装置を提供することができる効果がある。

【０１６４】また、請求項１１にかかる音声再生装置によれば、請求項１～８のいずれかに記載の音声再生装置において、音声復号化手段は、符号化音声信号を帯域毎に復号し、データ伸縮手段は、相関演算手段、帯域毎のクロスフェード手段を備え、相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、ピッチ周波数を包含する帯域においてセグメント間の相関値を演算し、相関値が高くなる位置で重み付け加算するためのシフト量を決定し、各クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段

で決定した位置で重み付け加算することを特徴とするものとしたので、フレームを構成するセグメント間の相関が高くなる位置に波形データをシフトさせて相関演算を行い、かつ各時間軸圧縮または時間軸伸長の処理において上記シフト量を考慮した処理を行うことによって、音声の基本周波数の周期性を保存するように、重み付け加算するフレームの位相の整合性を高められるため、音声信号の処理劣化の少ない速度変換処理を行うことができる音声再生装置を提供することができる効果がある。

【０１６５】また、請求項１２にかかる音声再生装置によれば、請求項１～８のいずれかに記載の音声再生装置において、音声復号化手段は、符号化音声信号を帯域毎に復号し、データ伸縮手段は、相関演算手段、帯域毎のクロスフェード手段を備え、相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、平均エネルギーが最大となる帯域においてセグメント間の相関値を演算し、相関値が高くなる位置で重み付け加算するためのシフト量を決定し、各クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算することを特徴とするものとしたので、フレームを構成するセグメント間の相関が高くなる位置に波形データをシフトさせて相関演算を行い、かつ各時間軸圧縮または時間軸伸長の処理において上記シフト量を考慮した処理を行うことによって、エネルギーが大きい主要な帯域での重み付け加算されるフレームの位相の整合性を高められるため、音声信号の処理劣化の少ない速度変換処理を行うことができる音声再生装置を提供することができる効果がある。

【０１６６】また、請求項１３にかかる音声再生装置によれば、請求項１～８のいずれかに記載の音声再生装置において、音声復号化手段は、符号化音声信号を帯域毎に復号し、データ伸縮手段は、相関演算手段、帯域毎のクロスフェード手段を備え、相関演算手段は、音声復号化手段から出力されるフレームを構成するセグメントの先頭位置を前回決定したシフト量に基づき補正し、各帯域においてセグメント間の相関値を演算し、相関値が最大の帯域において相関値が高くなる位置で重み付け加算するためのシフト量を決定し、各クロスフェード手段は、時間軸圧縮または時間軸伸長の際、音声復号化手段から出力されるフレームを構成するセグメントを、相関演算手段で決定した位置で重み付け加算することを特徴とするものとしたので、フレームを構成するセグメント間の相関が高くなる位置に波形データをシフトさせて相関演算を行い、かつ各時間軸圧縮または時間軸伸長の処理において上記シフト量を考慮した処理を行うことによって、最も周期性が存在し易いと予想される帯域での重み付け加算されるフレームの位相の整合性を高められるため、音声信号の処理劣化の少ない速度変換処理を行うことができる音声再生装置を提供することができる効果がある。

【０１６７】

【０１６８】

【０１６９】

【０１７０】

【０１７１】

【０１７２】

【０１７３】

【図面の簡単な説明】

【図１】本発明の実施の形態１による音声再生装置の全体ブロック図。

【図２】本発明の実施の形態１におけるデータ伸縮手段の構成図。

【図３】本発明の実施の形態１におけるデータ伸縮手段における一定値の時間軸圧縮／伸長の様子を示す　模式図。

【図４】本発明の実施の形態１における伸縮シーケンスの模式図。

【図５】本発明の実施の形態２におけるデータ伸縮手段の構成図。

【図６】本発明の実施の形態２におけるデータ圧縮の模式図。

【図７】本発明の実施の形態２におけるデータ圧縮の補正を行う場合のの模式図。

【図８】本発明の実施の形態２における他の例の伸縮シーケンスの模式図。

【図９】本発明の実施の形態３による音声再生装置の全体ブロック図。

【図１０】本発明の実施の形態４による音声再生装置のブロック図。

【図１１】本発明の実施の形態５による音声再生装置のブロック図。

【図１２】本発明の実施の形態６による音声再生装置のブロック図。

【図１３】本発明の実施の形態７による音声再生装置のブロック図。

【図１４】本発明の実施の形態８による音声再生装置のブロック図。

【図１５】本発明の実施の形態８における，エネルギー演算手段12-1-1がフレームのエネルギーを推定する過程を示すフローチャートである。

【図１６】本発明の実施の形態９による音声再生装置のブロック図。

【図１７】本発明の実施の形態１０による音声再生装置のブロック図。

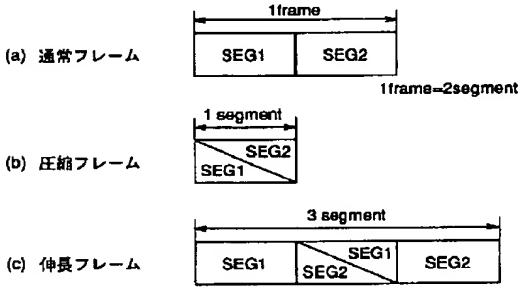
【図１８】本発明の実施の形態１１による音声再生装置のブロック図。

【図１９】従来の音声再生装置のブロック図。

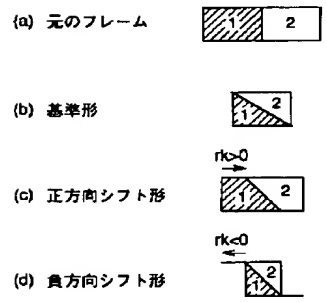
【図２０】従来の他の例の音声装置のブロック図。

【図２１】音声信号の主要ピッチ成分が含まれる周波数帯域について、その１フレーム分の時間軸波形を表した図。
【図２２】図２１に示した１フレームの信号を、その前半の信号部分と、後半の信号部分との２セグメントに分割して上下に並べた図。
【図２３】図２２における２セグメント間の相関関数を求めた値を示したグラフ。
【図２４】相関関数が最大となる時刻に後半の信号成分であるセグメントをずらせた様子を定性的に示した図。
【図２５】２セグメント間をTc時間オーバーラップさせてクロスフェード処理する様子を示した図。
【図２６】MPEG１オーディオレイヤ２の構成を示すブロック図。
【符号の説明】
１０１　フレーム逆パッキング手段
１０２　逆量子化手段
１０３　データ伸縮手段
１０４　サブバンド合成フィルタ手段
１０５　選択手段
１０６　フレームカウント手段
１０７　データ伸縮制御手段
１０８　フレームシーケンステーブル
２０１　バッファメモリ
２０２　クロスフェード手段
２０３　データ選択手段
３０１　相関演算手段
３０２　位相制御記憶手段
３００１　フレーム複号化手段
３００２　データ伸縮手段
３００３　伸縮頻度制御手段
３００４　エネルギー演算手段
３００５　フレーム選択手段
３００６　データ伸縮制御手段
４００４　音声らしさ演算手段
４００５　フレーム選択手段
５００４　定常性演算手段
５００５　フレーム選択手段
６００４　エネルギー変化度合演算手段
６００５　フレーム選択手段
７００５　フレーム選択手段

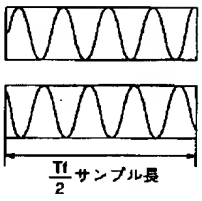
【図３】



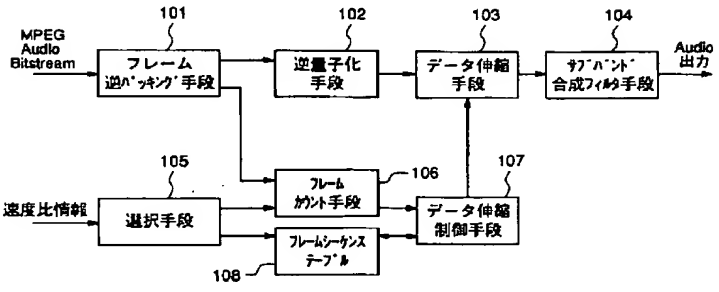
【図６】



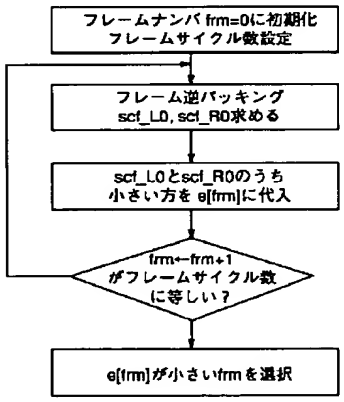
【図２２】



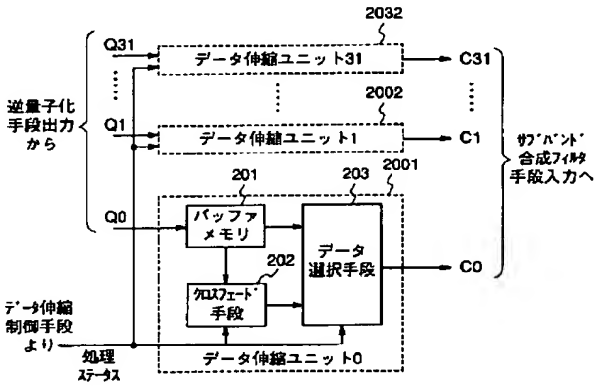
【図１】



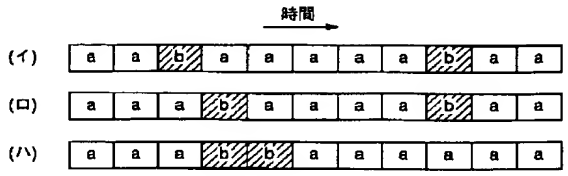
【図１５】



【図２】

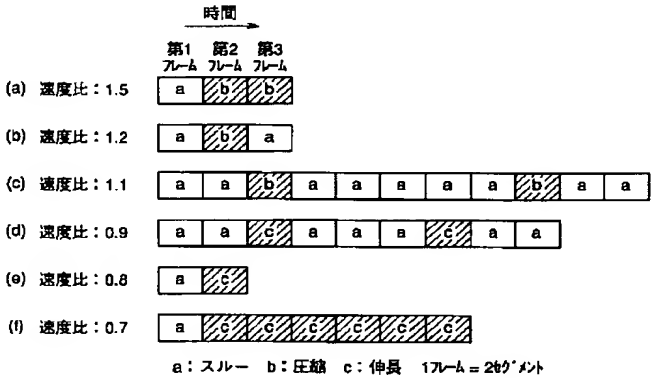


【図８】

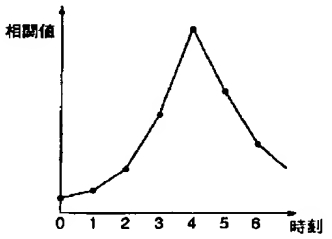


【図２１】

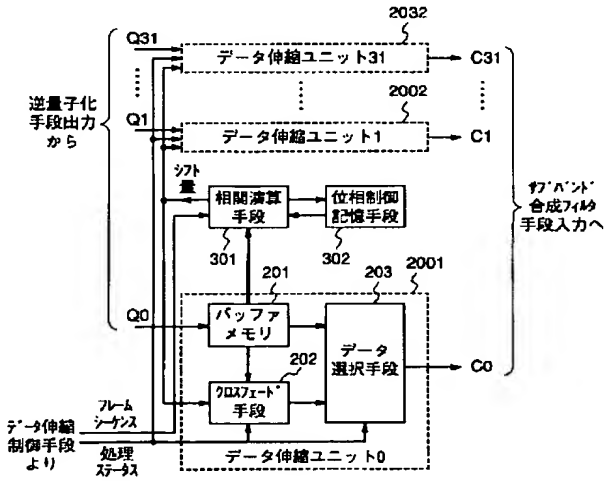
【図４】



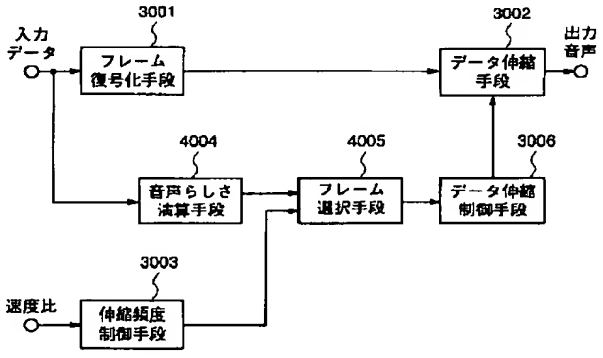
【図２３】



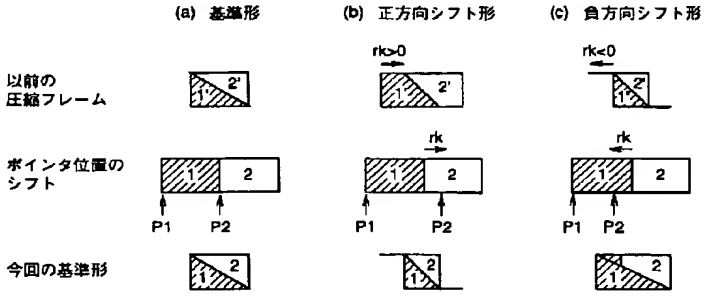
【図５】



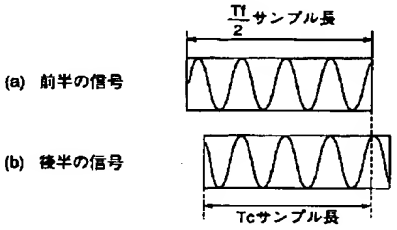
【図１０】



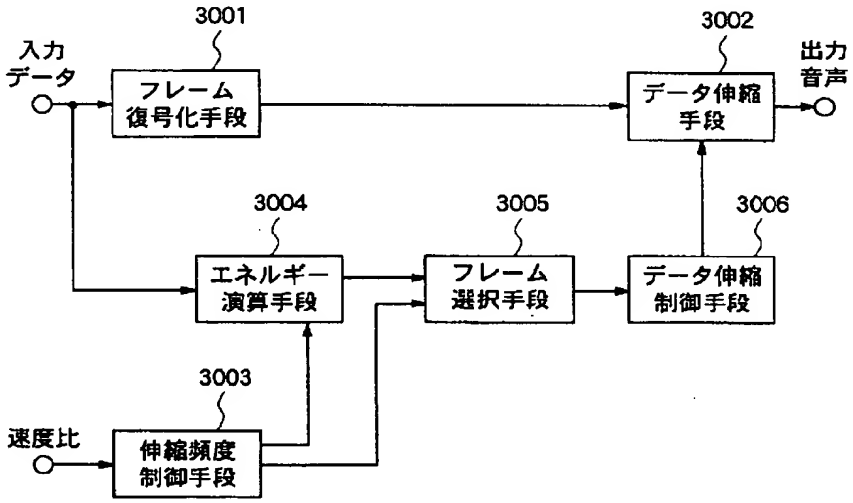
【図７】



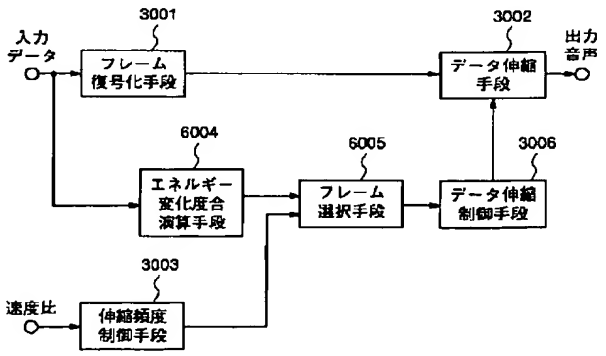
【図２４】



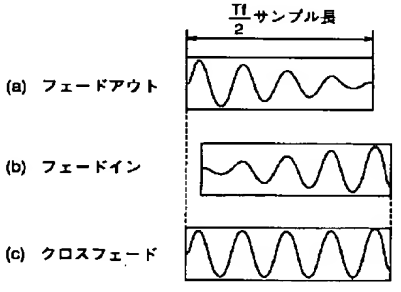
【図９】



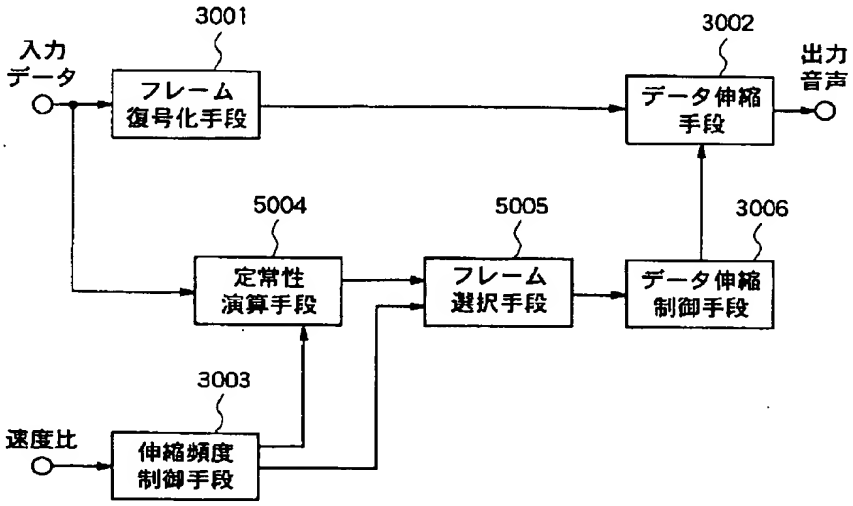
【図１２】



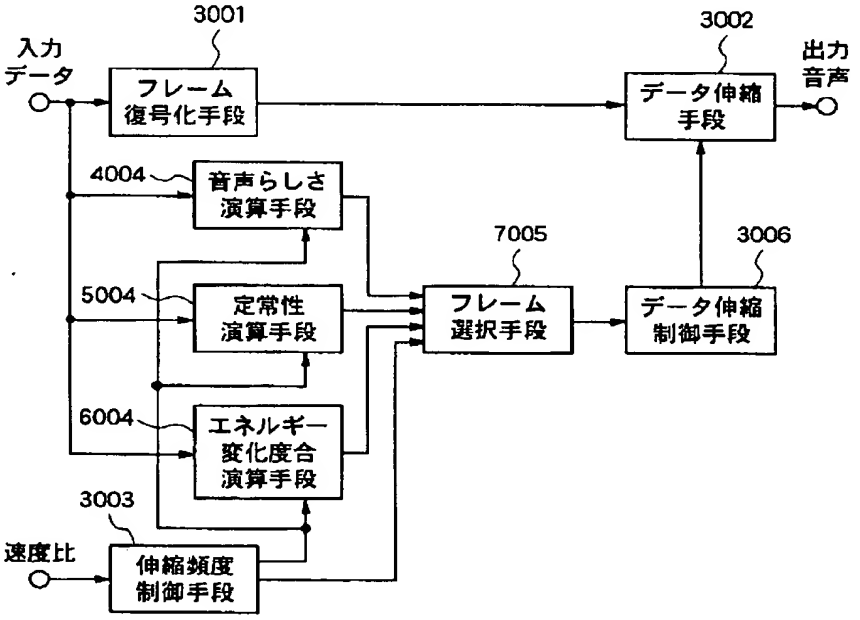
【図２５】



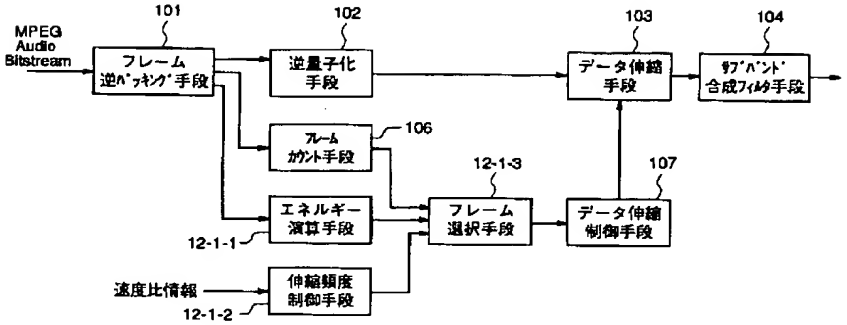
【図１１】


3001
入力
データ
フレーム
復号化手段
3002
データ伸縮
手段
出力
音声
5004
定常性
演算手段
5005
フレーム
選択手段
3006
データ伸縮
制御手段
3003
速度比
伸縮頻度
制御手段

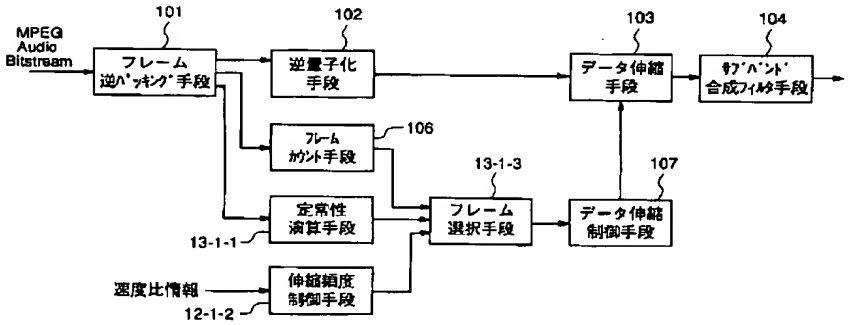
【図１３】


3001
入力
データ
フレーム
復号化手段
3002
データ伸縮
手段
出力
音声
4004
音声らしさ
演算手段
7005
3006
5004
定常性
演算手段
フレーム
選択手段
データ伸縮
制御手段
6004
エネルギー
変化度合
演算手段
3003
速度比
伸縮頻度
制御手段

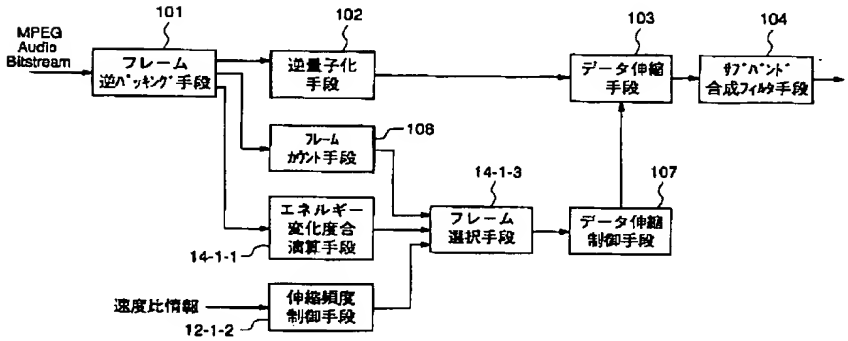
【図１４】


MPEG
Audio
Bitstream
101
フレーム
逆ﾊﾟｯｷﾝｸﾞ手段
102
逆量子化
手段
103
データ伸縮
手段
104
ｻﾌﾞﾊﾞﾝﾄﾞ
合成ﾌｨﾙﾀ手段
106
ﾌﾚｰﾑ
ｶｳﾝﾄ手段
12-1-3
107
エネルギー
演算手段
フレーム
選択手段
データ伸縮
制御手段
12-1-1
遠度比情報
伸縮頻度
制御手段
12-1-2

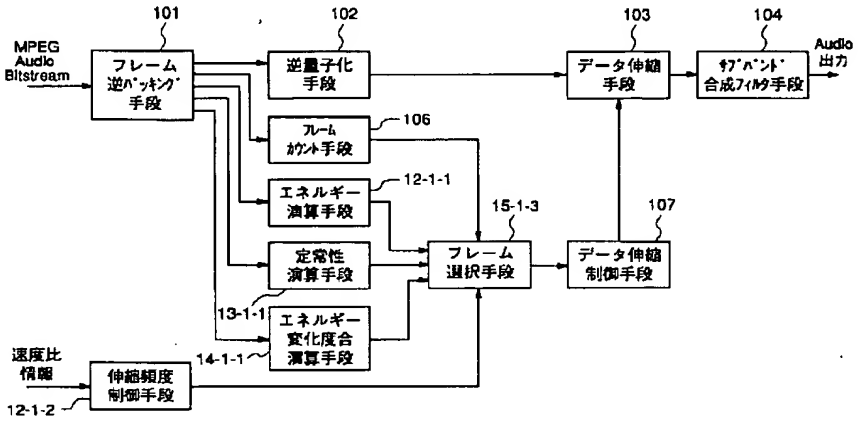
【図１６】


MPEG
Audio
Bitstream
101
フレーム
逆ﾊﾟｯｷﾝｸﾞ手段
102
逆量子化
手段
103
データ伸縮
手段
104
ｻﾌﾞﾊﾞﾝﾄﾞ
合成ﾌｨﾙﾀ手段
106
ﾌﾚｰﾑ
ｶｳﾝﾄ手段
13-1-3
107
定常性
演算手段
フレーム
選択手段
データ伸縮
制御手段
13-1-1
速度比情報
伸縮頻度
制御手段
12-1-2

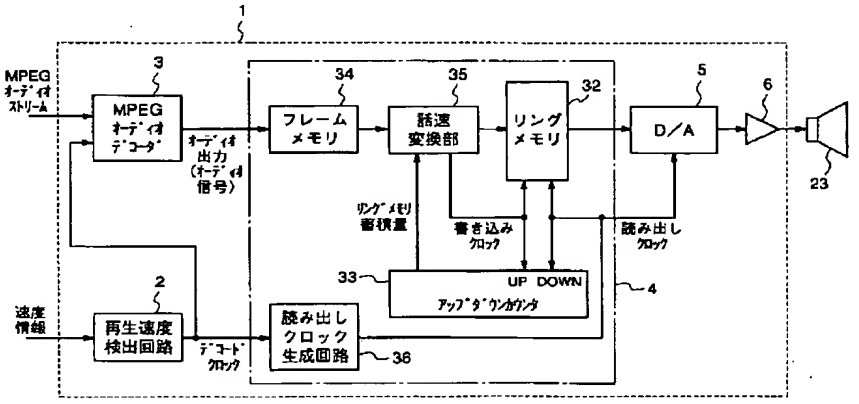
【図１７】


MPEG
Audio
Bitstream
101
フレーム
逆ﾊﾟｯｷﾝｸﾞ手段
102
逆量子化
手段
103
データ伸縮
手段
104
ｻﾌﾞﾊﾞﾝﾄﾞ
合成ﾌｨﾙﾀ手段
108
ﾌﾚｰﾑ
ｶｳﾝﾄ手段
14-1-3
107
エネルギー
変化度合
演算手段
フレーム
選択手段
データ伸縮
制御手段
14-1-1
速度比情報
伸縮頻度
制御手段
12-1-2

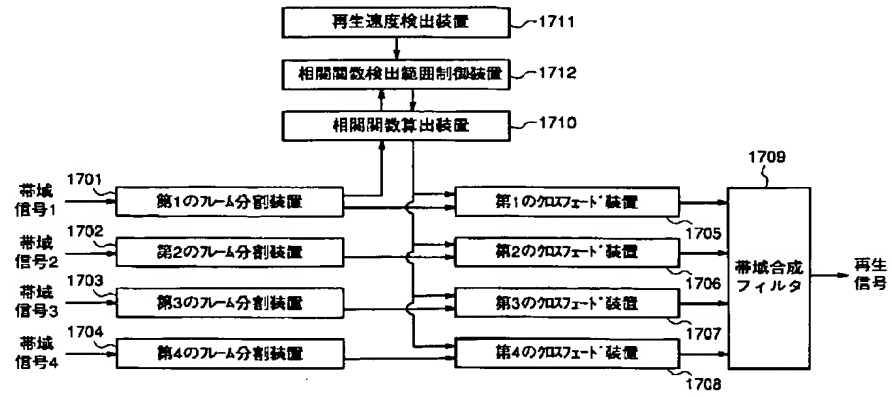
【図１８】


MPEG Audio Bitstream
フレーム 逆パッキング 手段
逆量子化 手段
データ伸縮 手段
サブバンド 合成フィルタ手段
Audio 出力
フレーム カウント手段
エネルギー 演算手段
定常性 演算手段
エネルギー 変化度合 演算手段
フレーム 選択手段
データ伸縮 制御手段
伸縮頻度 制御手段
速度比 情報
101
102
103
104
106
12-1-1
13-1-1
14-1-1
15-1-3
107
12-1-2

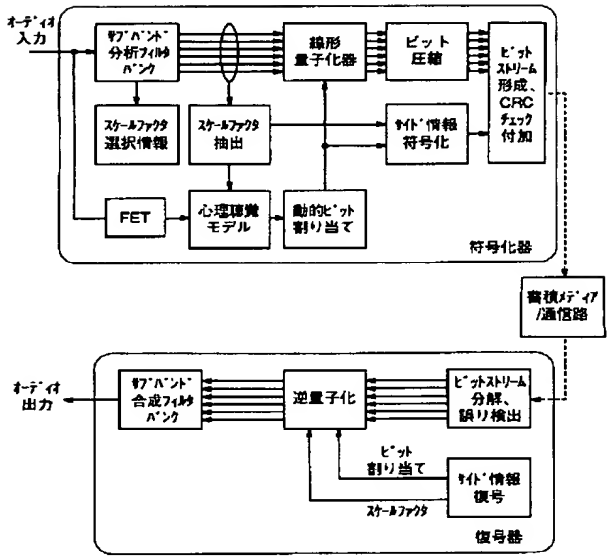
【図１９】


1
MPEG オーディオ ストリーム
3
MPEG オーディオ デコーダ
オーディオ 出力 (オーディオ 信号)
フレーム メモリ
34
話速 変換部
35
リング メモリ
32
D/A
5
6
23
リングメモリ 蓄積量
書き込み クロック
読み出し クロック
33
UP DOWN
アップダウンカウンタ
4
速度 情報
2
再生速度 検出回路
デコード クロック
読み出し クロック 生成回路
36

【図２０】


再生速度検出装置
1711
相関関数検出範囲制御装置
1712
相関関数算出装置
1710
1709
帯域 信号1
1701
第1のフレーム分割装置
第1のクロスフェード装置
1705
帯域 信号2
1702
第2のフレーム分割装置
第2のクロスフェード装置
1706
帯域 信号3
1703
第3のフレーム分割装置
第3のクロスフェード装置
1707
帯域 信号4
1704
第4のフレーム分割装置
第4のクロスフェード装置
1708
帯域合成 フィルタ
再生 信号

【図２６】



フロントページの続き

(72)発明者　松本　美治男
大阪府門真市大字門真1006番地　松下電器産業株式会社内

(56)参考文献　特開　平６－86164（ＪＰ，Ａ）
特開　平９－198088（ＪＰ，Ａ）
特開　平８－54895（ＪＰ，Ａ）
特開　平６－202692（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.7，ＤＢ名)
G10L　21/04
G10L　19/02
G11B　20/02